TEXAS A&M UNIVERSITY

COLLEGE STATION, TEXAS 77843-3143

Department of
STATISTICS
Phone 409 - 845-3141

DTIC FILE COPY

CONSIDERATIONS FOR THE LINEAR ESTIMATION
OF A REGRESSION FUNCTION
WHEN THE DATA ARE CORRELATED

D. B. Holiday, T. E. Wehrly and Jeffrey D. Hart

Department of Statistics
Texas A&M University

Technical Report No. 3

June 1987

DTIC
ELECTE
JUL 0 9 1987
S
E
D

87

kernel estimators is carried out. The expected values of the

estimators are compared against one another and against the optimal

linear combination that minimizes a certain measure of error.

Key Words: Nonparametric regression; growth curves; correlated data;

optimum bandwidth; mean averaged squared error.

Accession For		
NTIS GRA&I	☒	
DTIC TAB	☐	
Unannounced	☐	
Justification		
By		
Distribution/		
Availability Codes		
Dist	Avail and/or Special	
A-1		

Considerations for the Linear Estimation of a Regression Function When the Data are Correlated

D. B. Holiday, T. E. Wehrly and J. D. Hart

TEXAS A&M UNIVERSITY

Abstract

In fixed-design kernel nonparametric regression, there has been a paucity of results for models which allow for correlated errors. Consider the following repeated-measurements model, applicable in growth curve analysis: $Y_s(x_t) = g(x_t) + \epsilon_s(x_t)$, $s=1,\ldots,m$ (e.g.,subjects), $t=1,\ldots,n$ (e.g.,time points) with errors of zero mean and within-subject covariance matrix Σ. More specifically, we assume that $\mathrm{cov}[\epsilon_s(x_t),\epsilon_u(x_v)] = \delta_{su}\sigma(x_t,x_v)$ where δ_{su} is the Kronecker-delta and $\sigma(x_t,x_v)$ is the $(t,v)^{\text{th}}$ element of Σ. Furthermore, it is assumed that $\sigma(x_t,x_v)$ may be represented as the product of a scalar variance term and a suitably restricted correlation function $\gamma(x_t-x_v)$. Kernel estimators of the population regression function $g(x)$ are examined for specific as well as more general correlation functions. Limiting forms of an optimal linear combination of the subject means (and its measure of error) are derived. Necessary and sufficient conditions for consistency are stated for a general linear estimator for the Ornstein-Uhlenbeck correlation function, and sufficient conditions are given for a more general covariance structure. A numerical study investigating the requisite amount of smoothing and the efficiency of four popular

1. INTRODUCTION

1.1 The Model

Nonparametric regression has received considerable attention in the literature in the last several years. The usual model in the fixed-design case is

$$Y_i = g(x_i) + \epsilon_i, \quad i=1,\ldots,n; \tag{1.1.1}$$

$$E(\epsilon_i)=0, \quad Var(\epsilon_i) = \sigma^2 < \infty; \tag{1.1.2}$$

$$\epsilon_1,\ldots,\epsilon_n \text{ uncorrelated}; \tag{1.1.3}$$

$$x_1,\ldots,x_n \text{ selected by the experimenter.} \tag{1.1.4}$$

The goal is the estimation of the unknown regression function $g(x)$ from a sample $(x_1,Y_1),\ldots,(x_n,Y_n)$. More recently, the estimation of $g^{(p)}(x)$, the pth derivative of $g(x)$, has also become an important consideration, especially in growth curve applications. The term nonparametric derives from the lack of a finite-dimensional parameterization of the regression function and not from any distribution-free assumption on the part of the error terms. We will not be concerned with distributional assumptions and tests of hypotheses, the usual fare for parametric models. Certain smoothness constraints will be the only assumptions made for g.

Our aim is to investigate the estimation of $g^{(p)}(x)$, $p=0,1,2,\ldots$ in a model that relaxes the assumption of uncorrelated errors. Such a

model has been investigated by relatively few researchers. If the correlation matrix of the errors is known, the problem of estimating g would be greatly simplified. Without some method of estimating an unknown correlation structure (independent of the estimate of g), it would be difficult to assess how much of the data smoothness is due to g and how much is due to correlation. In a repeated-measurements model, however, one would possess enough observations to estimate the correlation independently of the estimate of g. We will therefore in this paper investigate a model which lends itself to such analysis:

$$Y_s(x_t) = g(x_t) + \epsilon_s(x_t); \qquad (1.1.5)$$

$$s = 1,2,\ldots,m, \quad t = 1,2,\ldots,n;$$

$$E[\epsilon_s(x_t)] = 0, \text{ all } s,t;$$

$$cov[\epsilon_s(x_t),\epsilon_u(x_v)] = \delta_{su}\sigma(x_t,x_v).$$

The quantity δ_{su} is the Kronecker-delta and $\sigma(x_t,x_v)$ is the covariance of errors with the same initial index (s=u). This is the type of model which would naturally arise in the growth-curve setting, an important problem in biological applications (see, for example, Grizzle and Allen,1969 and Morrison,1970). Consider a sample of m subjects, behaving independently, and measured for response Y at times $x_1,\ldots,x_n$. The regression function g(x) would represent the mean of the entire population of subjects from which we have a random sample of size m. The first and second derivatives of g(x) are of interest since they represent the velocity and acceleration,

respectively, of population growth. Since it is likely that the error term for a subject measured at time x_t is (positively) correlated with the error term at time x_{t+1}, we will have a set of correlated errors for that particular subject. The fact that the subjects are behaving independently would lead to an assumption of zero correlation across subjects. Most approaches assume that the subjects possess the same within-subject correlation structure. We will assume a certain form for the covariance term $\sigma(x_t, x_v)$, which in effect will be equivalent to homoscedastic and stationary errors. In particular, it will be assumed that the correlation may be written as a function of the difference in x_t and x_v. Homoscedasticity simply means that the variance of errors is constant over all points x in the range of interest. Stationarity implies that the correlation of the errors at different points only depends on the absolute distance between the points and not on their particular location. In the next chapter we will spell out the model in greater detail.

1.2   The Estimators

In the traditional model (1.1.1) there have been various approaches to the nonparametric estimation of g. Many evolved from the estimation of the conditional regression function

$$g^*(x) = E(Y|X=x). \qquad (1.2.1)$$

In this setting the design points $x_1, \ldots, x_n$ are not selected prior to the collection of the data, as in the fixed-design model. Instead, we

have a random sample $(X_1, Y_1), \ldots, (X_n, Y_n)$ from a bivariate

distribution with density $f(x,y)$, say. Let $f_X(x)$ denote the marginal

distribution of X, that is,

$$f_X(x) = \int_{-\infty}^{\infty} f(x,y)\, dy. \tag{1.2.2}$$

Then the conditional density of Y is

$$f_{Y|X}(y|x) = f(x,y)/f_X(x), \tag{1.2.3}$$

provided $f_X(x) > 0$. The conditional mean (1.2.1) -- the regression of

Y on X -- is then

$$g^*(x) = \int_{-\infty}^{\infty} y\, f_{Y|X}(y|x)\, dy = \int_{-\infty}^{\infty} y\, [f(x,y)/f_X(x)]\, dy. \tag{1.2.4}$$

Watson (1964) was the first to realize the connection between density

estimation and the functional (1.2.4). The estimator

$$g_n^{NW}(x) = \Sigma_{i=1}^{n} Y_i\, K[(x-x_i)/h] \,/\, \Sigma_{t=1}^{n} K[(x-x_t)/h] \tag{1.2.5}$$

was also independently proposed by Nadaraya (1964) and hence is known

as the Nadaraya-Watson (NW) estimator. The function K is known as a

kernel and was originally restricted to be a probability density

function. The quantity h is known as the bandwidth and controls the

amount of smoothing done by the estimator. A large bandwidth gives

greater influence (weight) to Y values corresponding to X values less

local to the point of estimation x, thereby resulting in a smoother

estimate. The value of $h > 0$ and the kernel function K are user-

supplied.

Kernel methodology has its roots in the nonparametric estimation

of an unknown continuous probability density function, $f(x)$, given

data $X_1, \ldots, X_n$. No paper on nonparametric estimation of functions

would be complete without reference to the classic papers of

Parzen (1962a) and Rosenblatt (1956), which form the basis of kernel estimation.

In the fixed-design model (1.1.1), Priestley and Chao (1972) were the first to propose the following kernel estimate, which bears their name:

$$g_n^{PC}(x) = (1/h) \sum_{i=1}^n y_i (x_{i+1} - x_i) K[(x - x_i)/h]. \qquad (1.2.6)$$

This estimate has been widely studied, as will be seen in the references of Section 1.5. Many variations of this estimator have been proposed, most notably the Gasser-Müller (GM) estimate (Gasser and Müller, 1979). These authors, using an argument based on the mean value theorem, propose

$$g_n^{GM}(x) = \sum_{i=1}^n y_i \left\{ (1/h) \int_{A_i} K[(x-u)/h] du \right\}, \qquad (1.2.7)$$

where $A_i = [s_i, s_{i+1}]$ with the s_i-values chosen to satisfy $x_i \leq s_i \leq x_{i+1}$. Notice that these estimators are all of the linear form

$$g_n(x) = \sum_{i=1}^n w_{ni}(x) y_i \qquad (1.2.8)$$

for appropriate choices of $w_{ni}(x)$. Also observe that the weights in (1.2.5) sum to one, and therefore the NW estimate is a true weighted average. Even though originally proposed for conditional mean estimation, the NW estimate is commonly used in the fixed design model. If the design points are equally spaced, the NW estimate may be obtained from the PC estimate by dividing by the sum of the weights in (1.2.6). Such estimators are said to be "cut-and-normalized", and often reduce the bias of the estimator, especially if predicting at an x near the boundary of the range of interest.

In this paper we will consider only properties of an optimal linear estimate (1.2.8) and of kernel-based estimates along with their cut-and-normalized versions for equally-spaced design points. With the addition of the cut-and-normalized GM estimate, which we will denote GC, there are then four kernel estimates: PC, GM, NW, GC. When estimating $g(x)$ in our repeated-measurements model, the y_t-values will be replaced by the subject means $\bar{y}(x_t)$, $t=1,\ldots,n$.

Other methods of nonparametric estimation of the regression function which have attained stature include the method of stochastic approximation, nearest-neighbor estimation, the regressogram, and the chief competitor of kernel estimation -- splines. The scope of this paper, however, is restricted to kernel estimation. The reader is referred to Prakasa Rao (1983) for a brief summary of the above methods.

1.3 The Problem

Given the repeated-measurements model (1.1.5), the foremost aim of this research is to assess the performance of the four kernel estimators PC, GM, NW, and GC when the only allowance made for correlation is bandwidth adjustment. Pursuant to this goal, the estimators will be compared with one another and with some standard which represents the best possible linear estimator one could use. Of course, we will have to define what is meant by the term "best". The loss in performance of the kernel estimators for a given variance and

correlation situation may point to a need to incorporate some
estimate of the correlation into the estimator (a la generalized
least squares), or, to possibly abandon kernel estimation completely
in favor of one of the competing methods mentioned earlier. We will
therefore need to obtain and investigate some optimal linear
combination of the subject means to afford a standard for
comparisons. Limiting expressions of this linear estimator will shed
light onto consistency properties as well as suggest a possible
modification of the kernel function. We will deal with these goals in
Chapter 2, which will focus on general linear estimation -- not
restricted to kernel form.

Chapter 3 will review important results in kernel estimation of
$g^{(0)}(x)$, including the recent work of Hart and Wehrly (1986). These
authors consider the same repeated-measurements model and develop
results primarily for the GM estimator under fairly general
correlation structures. Chapter 3 will also describe the results of a
numerical study designed to measure the efficacy of kernel estimates
relative to the best linear combination. A variety of functions and
variance-correlation settings will be considered. The uncorrelated
case will be included for comparison. The optimal amount of smoothing
selected by the estimators (according to minimization of a measure of
global error) will be a by-product of the study and may be used to
study bandwidth variation as a function of the amount of variance and
correlation present in the model.

1.4 Some Theoretical Terms and Definitions

For the benefit of the reader who is unfamiliar with the nature of the research being done in this area, we will give a sampling of technical definitions which are common in the literature, which is reviewed in the subsequent section. Many of the results pertain to verification of certain asymptotic (large sample) properties for various estimators. These include probability bounds on the error of estimation, asymptotic normality, and various modes of consistency, many of which are given below.

Let $g_n(x)$ denote an estimator of $g(x)$ in a model similar to (1.1.1). We first define two criteria for nearness of the estimator to the function:

<u>Definition (1.4.1)</u> The quantity

$$MSE[g_n(x)] = E[g_n(x)-g(x)]^2 \qquad (1.4.1)$$

is known as the <u>mean squared error</u> of g_n at the point x.

<u>Definition (1.4.2)</u> The quantity

$$MISE[g_n,g] = E \int [g_n(x)-g(x)]^2 dx = \int E[g_n(x)-g(x)]^2 dx \qquad (1.4.2)$$

is known as the <u>mean integrated squared error</u> of g_n with respect to g.

Letting G^* denote a space of functions, the following modes of convergence are defined for a sequence g_n of estimators:

<u>Definition (1.4.3)</u> Consistency in Quadratic Mean.

For every x and $g \in G^*$, $E[g_n(x)-g(x)]^2 \rightarrow 0$, as $n \rightarrow \infty$.   (1.4.3)

<u>Definition (1.4.4)</u> Integratedly Consistent in Quadratic Mean.

For every $g \in G^*$, $E\int[g_n(x)-g(x)]^2dx \rightarrow 0$, as $n \rightarrow \infty$.   (1.4.4)

<u>Definition (1.4.5)</u> Weakly Consistent.

For every x and $g \in G^*$, $g_n(x) \xrightarrow{P} g(x)$, as $n \rightarrow \infty$.   (1.4.5)

<u>Definition (1.4.6)</u> Uniformly Weakly Consistent.

For every $g \in G^*$, $\sup_x |g_n(x) - g(x)| \xrightarrow{P} 0$, as $n \rightarrow \infty$.   (1.4.6)

<u>Definition (1.4.7)</u> Strongly Consistent.

For every x and $g \in G^*$, $g_n(x) \xrightarrow{wp1} g(x)$, as $n \rightarrow \infty$.   (1.4.7)

<u>Definition (1.4.8)</u> Uniformly Strongly Consistent.

For every $g \in G^*$, $\sup_x |g_n(x) - g(x)| \xrightarrow{wp1} 0$, as $n \rightarrow \infty$.   (1.4.8)

<u>Definition (1.4.9)</u> Asymptotically Unbiased.

For every x and $g \in G^*$, $E[g_n(x)] \rightarrow g(x)$, as $n \rightarrow \infty$.   (1.4.9)

<u>Definition (1.4.10)</u> Uniformly Asymptotically Unbiased.

For every $g \epsilon G^*$, $\sup_x |E[g_n(x)] - g(x)| \rightarrow 0$, as $n \rightarrow \infty$. (1.4.10)

As in density estimation, there exist classes of functions for which unbiased estimators do not exist and for which uniformly consistent estimators do not exist. The reader is referred to Prakasa Rao (1983) for more details.

We will deal primarily with pointwise consistency in mean squared error (1.4.3) and global consistency in mean integrated squared error (1.4.4). Since the n-fold integral in (1.4.4) presents analytic and computational problems, we will later introduce a discrete analog known as the mean <u>averaged</u> squared error.

1.5 A Review of Selected Literature

The literature on nonparametric regression is presently developing at such a rate that the body of knowledge will soon rival that of nonparametric density estimation. Of course, this is to be expected since most nonparametric methods are adapted and motivated from existing density estimation methodology. For the reader not familiar with density estimation, Tapia and Thompson (1978) contains a useful survey. Prakasa Rao (1983) not only provides a comprehensive review of density estimation, but also reviews estimation of functionals related to densities, which includes the regression function. Collomb (1981, 1985) and Stone (1977) also provide extensive

bibliographies on nonparametric regression. The focus of the latter paper is on estimation in the random regressor model (1.2.4). An excellent and up-to-date survey of fixed-design methods (particularly for splines) may be found in Eubank (1986).

The Nadaraya-Watson estimator is usually considered in the context of the random regressor model. The following authors have considered this estimator and its variations: Rosenblatt (1969), Nadaraya (1970, 1973, 1983a, 1983b), Schuster (1972), Noda (1976), Devroye (1978), Schuster and Yakowitz (1979), Devroye and Wagner (1980a, 1980b), Spiegelman and Sacks (1980), Greblicki and Krzyzak (1980), Mack and Silverman (1982), Johnston (1982), and others.

As previously mentioned, the Gasser-Müller estimator is a variation of the Priestley-Chao estimator. Authors considering forms of these estimators include Benedetti (1974, 1975, 1977), Clark (1977), Gasser and Müller (1979), Cheng and Lin (1981a, 1981b), Georgiev (1984a, 1984b, 1984c, 1984d), Georgiev and Greblicki (1986), and others. The last paper also contains useful results for arbitrary linear estimators.

Gasser, Müller, Köhler, Molinari, and Prader (1984) consider applications of kernel methodology, including derivative estimation, to the estimation of individual growth-curves. Azzalini (1984) and Glasbey (1979) attack the growth-curve problem from a parametric point of view. A linear model with autocorrelated errors is considered by the former author while the latter author studies

estimation in a nonlinear regression model, also with an autocorrelated error structure similar to ours. Diggle and Hutchinson (1985) consider (nonparametric) spline estimation under an autocorrelated errors model. An approach based on regression analysis of a continuous parameter time series is considered in several papers found in Parzen (1967).

2. OPTIMAL LINEAR ESTIMATION

2.1 Model and Estimator Notation

As mentioned in Chapter 1, the development of an optimal linear combination of the data, not necessarily of kernel form, is now considered. First the model and the assumptions are explicitly stated:

$$Y_s(x_t) = g(x_t) + \epsilon_s(x_t);$$

$$s = 1,2,\ldots,m \qquad t = 1,2,\ldots,n;$$

$$E[\epsilon_s(x_t)] = 0, \text{ all } s,t;$$

$$\text{cov}[\epsilon_s(x_t),\epsilon_u(x_v)] = \delta_{su}\sigma(x_t,x_v). \tag{2.1.1}$$

Let the vector of design points and corresponding true function values be denoted by

$$X = [x_1,\ldots,x_n]' \text{ and} \tag{2.1.2}$$

$$g = [g(x_1),\ldots,g(x_n)]'. \tag{2.1.3}$$

When it suits our purpose, we will often use x_i and x_{ni} interchangeably to emphasize the dependence of a sequence of design points on n. In our development we will usually use an equally-spaced set of design points on [0,1], namely,

$$x_{ni} = (i-1)/n, \tag{2.1.4}$$

which has spacing $(1/n)$ and often simplifies the arguments.

Let Y and ϵ denote the (m by n) data and error matrices, respectively. In matrix notation, the model (2.1.1) may be expressed

$$Y = 1_m g' + \epsilon, \quad E[\epsilon] = 0, \quad \text{cov}[\text{vec}(\epsilon')] = I_m \otimes \Sigma, \quad (2.1.5)$$

where 1_m is an m-vector of 1's, vec is the stacking operator, and $\otimes$ denotes the Kronecker product. Note that the m rows of ϵ are uncorrelated with each other. Each row has (n by n) covariance structure Σ, whose $(i,j)^{\text{th}}$ element is $\sigma(x_i, x_j)$. In particular we will be interested in the case when $\Sigma = \sigma^2 \Gamma$, with the $(i,j)^{\text{th}}$ element of the correlation matrix Γ being $\gamma(x_i - x_j)$. Note that γ is an arbitrary correlation function assumed to satisfy certain properties as the need arises. The specific example we will use is the Ornstein-Uhlenbeck structure

$$\gamma(u) = \exp\{-\alpha|u|\}, \quad \alpha > 0, \quad (2.1.6)$$

which for $\rho > 0$ is a reparameterization of

$$\gamma(u) = \rho^{|u|}, \quad 0 < \rho < 1. \quad (2.1.7)$$

Therefore the uncorrelated case ($\rho = 0$) corresponds to $\alpha \rightarrow \infty$ and the unit correlation ($\rho = 1$) corresponds to $\alpha \rightarrow 0$.

For estimation of the regression function $g(x)$ at the point x, we will entertain the estimator

$$g_n(x) = \Sigma_{i=1}^n w_{ni}(x) \, \bar{y}(x_i), \quad (2.1.8)$$

where

$$\bar{y}(x_i) = (1/m) \, \Sigma_{s=1}^m Y_s(x_i) \quad (2.1.9)$$

and $w_{ni}(x)$ is an arbitrary weight function which, in addition to dependence on x, may also depend on the entire design vector (2.1.2). As in (2.1.3), we denote

$$g_n = [g_n(x_1), \ldots, g_n(x_n)]' \qquad (2.1.10)$$

as the vector of estimated function values at the design points.
Notice that (2.1.8) is a special case of the more general form

$$g_n^o(x) = \Sigma_i \ \Sigma_s \ w_{ni}(x) \ d_{ms} \ Y_s(x_i) \qquad (2.1.11)$$

with $d_{ms} = 1/m$, $s = 1, 2, \ldots, m$. We will subsequently show that there
is no need to allow for the extra coefficients d_{ms} in considering our
global optimality criterion.

2.2 Mean Averaged Squared Error

As a global measure of discrepancy between the vectors g_n (the
estimated values at the design points) and g (the true values at the
design points), we adopt the mean averaged squared error,

$$MASE[g_n, g] = E\{(1/n) \ \Sigma_{i=1}^n \ [g_n(x_i) - g(x_i)]^2\}, \qquad (2.2.1)$$

which is an approximation of the mean integrated squared error,

$$MISE[g_n, g] = E \int [g_n(x) - g(x)]^2 dx. \qquad (2.2.2)$$

The MISE is dealt with in theoretical situations whereas it is much
more practical to use the MASE in computer approximations and data-
oriented applications. We should note in passing that

$$MASE[g_n, g] \sim MISE[g_n, g],$$

provided that we have an equally-spaced or asymptotically equally-
spaced design. To see this, let

$$F_n(x) = (1/n) \ \Sigma_i \ I\{x_i \leq x\},$$

where $I\{\cdot\}$ is an indicator function. Then $F_n(x)$ could be considered a
sample cumulative distribution function converging to

$$F(x) = \int_{-\infty}^{x} f(t)dt,$$

where $f(t)$ is called a <u>design density</u>. We may choose the design points as selected quantiles of this distribution or as a random sample from this distribution, in which case $MASE[g_n, g]$ may be written

$$E \int [g_n(x) - g(x)]^2 dF_n(x) \sim E \int [g_n(x) - g(x)]^2 f(x)dx, \text{ as } n \to \infty. \quad (2.2.3)$$

We see that the last expression in (2.2.3) defines a weighted version of the MISE. Recalling the design points (2.1.4), we see that this will lead to a uniform design density on $[0,1]$ and hence the right side of (2.2.3) corresponds to the usual uniformly-weighted MISE in (2.2.2).

Let W_n denote the matrix whose $(i,j)^{th}$ element is $w_{ni}(x_j)$. Then we may write g_n as

$$g_n = (1/m) W_n' Y' 1_m, \quad (2.2.4)$$

using notation of the previous section. The matrix equivalent of the general form (2.1.11) is

$$g_n^o = W_n' Y' d_m \quad (2.2.5)$$

where $d_m = [d_{m1}, \ldots, d_{mm}]'$.

We proceed with the main result of this section.

<u>Theorem (2.2.1)</u> Consider the model (2.1.1) and the general estimator (2.2.5). If our optimality criterion is the minimization of MASE subject to the constraint

$$1_m' d_m = 1, \quad (2.2.6)$$

then for $G = gg'$ and $\Sigma_m = (1/m)\Sigma$,

(i) $\quad M(W_n) \equiv \text{MASE}[g_n^o, g] = (1/n)\text{tr}\{(d_m'd_m)W_n'\Sigma W_n + (W_n-I)'G(W_n-I)\}$. (2.2.7)

(ii) The optimal choice of d_m is

$$d_m^* = (1/m)1_m.\qquad\qquad (2.2.8)$$

(iii) The optimal choice of W_n is

$$W_n^* = (\Sigma_m + G)^{-1}G = [1+g'\Sigma_m^{-1}g]^{-1}\,\Sigma_m^{-1}gg'.\qquad (2.2.9)$$

(iv) The minimal MASE, $M(W_n^*)$, is

$$[(g'g)/n][1 - \text{tr}(W_n^*)] = [(g'g)/n][1 - g'(\Sigma_m + G)^{-1}g].\ (2.2.10)$$

<u>Proof:</u> For (i) we note that $n \cdot \text{MASE}[g_n^o, g] = E||g_n^o - g||^2$

$$= E||W_n'Y'd_m - W_n'g1_m'd_m + W_n'g1_m'd_m - g||^2$$

$$= E||W_n'\epsilon'd_m + (W_n-I)'g||^2$$

$$= E||W_n'\epsilon'd_m||^2 + E||(W_n-I)'g||^2 \text{ (where cross products vanish)}$$

$$= E[\text{tr}\{d_m'\epsilon W_n W_n'\epsilon'd_m\}] + \text{tr}\{g'(W_n-I)(W_n-I)'g\}$$

$$= E[\text{tr}\{(W_n'\epsilon'd_m)(W_n'\epsilon'd_m)'\}] + \text{tr}\{(W_n-I)'gg'(W_n-I)\}$$

$$= \text{tr}\{\text{var}(W_n'\epsilon'd_m)\} + \text{tr}\{(W_n-I)'G(W_n-I)\}.$$

Recall that the rows of ϵ are uncorrelated and have common covariance matrix Σ. Let $\epsilon' = [\epsilon_1, \ldots, \epsilon_m]$ denote the m uncorrelated columns. Then

$$W_n'\epsilon'd_m = \Sigma_{s=1}^m W_n'\epsilon_s d_{ms},\qquad\qquad (2.2.11)$$

which is a linear combination of the uncorrelated quantities $W_n'\epsilon_s$. Hence

$$\text{var}[W_n'\epsilon'd_m] = \Sigma_{s=1}^m d_{ms}^2\text{var}[W_n'\epsilon_s] = (d_m'd_m)W_n'\Sigma W_n.\qquad (2.2.12)$$

Equation (2.2.7) immediately follows upon substitution into the last expression for MASE.

For (ii) it is widely known that the solution to the minimization problem

$$\min\ d_m'd_m$$

$$\text{s. t.}\quad 1_m'd_m = 1 \qquad\qquad (2.2.13)$$

is $d_m^* = (1/m)1_m$. Since this value is an attainable lower bound independent of W_n, this is also the solution to the original question, due to the nonnegativity of the other quantities in (2.2.7). Note that the first MASE term represents the contribution due to the variance and the last term represents the squared bias.

For (iii) we will substitute the optimal d_m (2.2.8) into (2.2.7), obtaining

$$\text{MASE}[\varphi_\lambda, g] = M(W_n) = (1/n)\,\text{tr}\{W_n'\Sigma_m W_n + (W_n - I)'G(W_n - I)\}. \quad (2.2.14)$$

Due to the symmetry of G and the properties of traces we may write this as

$$M(W_n) = (1/n)\,\text{tr}\{W_n'(\Sigma_m + G)W_n - 2W_n'G + G\}. \qquad (2.2.15)$$

To find the optimal value of W_n, we may proceed in either of two directions. We may use the properties of matrix calculus to differentiate the trace directly, or we may express the MASE as a quadratic form of a Kronecker product, and then take the familiar derivative. Following the former approach, we obtain

$$\partial M(W_n)/\partial W_n = [(\Sigma_m + G) + (\Sigma_m + G)']W_n - 2G. \qquad (2.2.16)$$

Upon setting this value equal to zero, the result immediately follows. Note that the fact that $(\Sigma_m + G)$ is positive definite will guarantee that (2.2.15) is indeed minimized. The alternate expression in (2.2.9) for W_n^* follows from the matrix fact

$$(I + AB)^{-1} = I - A(I + BA)^{-1}B, \qquad (2.2.17)$$

by factoring out $\Sigma_{\mathbf{m}}^{-1}$ and taking $A = g$ and $B = g'\Sigma_{\mathbf{m}}^{-1}$. The calculations are straightforward but tedious and hence omitted.

For (iv), the result is found easily by substituting (2.2.9) into (2.2.15) and using the well-known property $\text{tr}(AB) = \text{tr}(BA)$. ∎

In light of (2.2.8), we will only consider estimators of the form (2.1.8) and not of the more general form (2.1.11). We also note that if one computes $W_n^\bullet$, then the left side of (2.2.10) is an efficient formula for calculating the minimal MASE for a known regression function since it involves fewer operations than the other formulas.

2.3 General Conditions Sufficient for Consistency

Using (2.2.9), $M(W_n^\bullet)$, the measure of error evaluated at the optimum, may be expressed as

$$M(W_n^\bullet) = [(g'g)/n][1 - \text{tr}(W_n^\bullet)] = [(g'g)/n][1+g'\Sigma_{\mathbf{m}}^{-1}g]^{-1}. \quad (2.3.1)$$

By properties of Riemann sums,

$$[(g'g)/n] \longrightarrow \int g^2(x)dx, \text{ as } n \longrightarrow \infty, \qquad (2.3.2)$$

which is finite if g is continuous on $[0,1]$. From (2.3.1) we see that in order for $M(W_n^\bullet) \longrightarrow 0$ as $n \longrightarrow \infty$ (with m bounded), we will need

$$g'\Sigma_{\mathbf{m}}^{-1}g = (m/\sigma^2)g'\Gamma^{-1}g \longrightarrow \infty, \text{ as } n \longrightarrow \infty, \qquad (2.3.3)$$

unless m is allowed to tend to infinity.

In the following argument, let

$$e_{n1}(A) \geq \dots \geq e_{nn}(A) \qquad (2.3.4)$$

denote the descending eigenvalues of an (n by n) positive definite matrix A.

Theorem (2.3.1) Suppose g is continuous on $[0,1]$ and not identically 0. If $m \rightarrow \infty$ or if $n/e_{n1}(\Gamma) \rightarrow \infty$ as $n \rightarrow \infty$, then $M(W_n^*) \rightarrow 0$.

Proof: In light of (2.3.3) the result will follow by showing $g'\Sigma_m^{-1}g \rightarrow \infty$. Using the optimization lemmas for positive definite quadratic forms,

$$g'\Sigma_m^{-1}g = (m/\sigma^2)g'\Gamma^{-1}g \geq (m/\sigma^2)g'g\, e_{nn}(\Gamma^{-1}). \qquad (2.3.5)$$

By the spectral decomposition, the eigenvalues of Γ^{-1} are positive and are the reciprocals of the eigenvalues of Γ. Furthermore, $n/e_{n1}(\Gamma)$ is greater than one for all n since $e_{n1}(\Gamma) < \Sigma_{i=1}^{n} e_{ni}(\Gamma) = tr(\Gamma) = n$.

Hence

$$\sigma^2 g'\Sigma_m^{-1}g \geq mn[(g'g)/n]/e_{n1}(\Gamma) \qquad (2.3.6)$$

$$\sim \int g^2(x)dx\, (mn/e_{n1}(\Gamma)), \qquad (2.3.7)$$

from which the result clearly follows since the integral is strictly positive. ∎

In a subsequent section, the eigenvalue condition of this theorem will be related to the properties of the spectral density associated with the Ornstein-Uhlenbeck correlation function.

2.4 Results for the Ornstein-Uhlenbeck Structure

In this section we consider applications of the previous results as in the case of the Ornstein-Uhlenbeck covariance structure (2.1.6) in the equally-spaced case (2.1.4). Let

$$\rho = \exp(-a), \quad a > 0 \tag{2.4.1}$$

be a nonnegative constant and define

$$\rho_n = \rho^{1/n} = \exp(-a/n), \tag{2.4.2}$$

where $1/n$ is the spacing of the design points on $[0,1]$. In this case the correlation matrix Γ has only n distinct elements and has the symmetric Toeplitz form with $(i,j)^{th}$ element $\rho_n^{|i-j|}$, due to the fact that $|x_i - x_j| = |i-j|/n$. Since the elements of the first row of a symmetric Toeplitz matrix determine the distinct elements, we often use the special notation

$$\Gamma = STOEPL(1, \rho_n, \ldots, \rho_n^{n-1}) \tag{2.4.3}$$

to refer to this matrix. In general, a matrix A is said to be Toeplitz if its $(i,j)^{th}$ element is some function, say $a(j-i)$, as $j-i$ ranges from $1-n$ to $n-1$. We write this using notation

$$A = TOEPL[a(1-n), \ldots, a(0), \ldots, a(n-1)]. \tag{2.4.4}$$

Notice that STOEPL requires that $a(i) = a(-i)$.

We are interested in finding necessary and sufficient conditions under which $M(W_n^\bullet) \to 0$. The following lemma will be useful in this regard.

<u>Lemma (2.4.1)</u> Suppose

$$x_i = x_{ni} = (i-1)/n, \quad i=1,2,\ldots,n,$$

$$g = [g(x_{n1}),\ldots,g(x_{nn})]', \text{ and}$$

$$\Gamma = \text{STOEPL}(1,\rho_n^1,\ldots,\rho_n^{n-1}).$$

Then the quadratic form $g'\Gamma^{-1}g$ may be written

$$g'\Gamma^{-1}g = (1-\rho_n^2)^{-1} [B_{n1} + S_n + B_{nn}], \qquad (2.4.5)$$

where S_n, the sum of the interior terms, is

$$S_n = \Sigma_{i=2}^{n-1} \{(1+\rho_n^2)g^2(x_{ni}) - \rho_n g(x_{ni})[g(x_{n,i-1}) + g(x_{n,i+1})]\}, \qquad (2.4.6)$$

and B_{n1}, B_{nn} are boundary terms with

$$B_{n1} = g(x_{n1}) [g(x_{n1}) - \rho_n g(x_{n2})], \text{ and} \qquad (2.4.7)$$

$$B_{nn} = g(x_{nn}) [g(x_{nn}) - \rho_n g(x_{n,n-1})]. \qquad (2.4.8)$$

Furthermore, if g has at least two continuous derivatives, is right

and left differentiable at 0 and 1, respectively, and $g^2(x)$,

$g(x)g'(x)$ are integrable on $[0,1]$, then

$$(1-\rho_n^2)^{-1} S_n \longrightarrow (1/2a) \int_0^1 g(x)[a^2g(x)-g'(x)]dx, \qquad (2.4.9)$$

$$(1-\rho_n^2)^{-1} B_{n1} \longrightarrow (1/2a) g(0)[ag(0)-g'(0+)], \text{ and} \qquad (2.4.10)$$

$$(1-\rho_n^2)^{-1} B_{nn} \longrightarrow (1/2a) g(1)[ag(1)+g'(1-)]. \qquad (2.4.11)$$

<u>Proof:</u> The inverse of Γ is a patterned matrix of tridiagonal form.

Direct multiplication yields the form (2.4.5) for $g'\Gamma^{-1}g$.

Next we factor out $\rho_n = e^{-a/n}$ in S_n, obtaining

$$S_n = e^{-a/n} \Sigma_{i=2}^{n-1} g(x_{ni})\{[e^{a/n}+e^{-a/n}]g(x_{ni})-[g(x_{n,i-1})+g(x_{n,i+1})]\}. \ (2.4.12)$$

Now, adding the Taylor series expansions for $e^{a/n}$ and $e^{-a/n}$ yields

$$2 + (a^2/n^2) + O(n^{-4}). \qquad (2.4.13)$$

Expanding $g(x_{n,i-1})$ and $g(x_{n,i+1})$ about the point x_{ni} yields

$$g(x_{n,i-1}) = g(x_{ni}) + g'(x_{ni})(-1/n) + g''(x_{ni}^o)(1/n)^2/2!, \text{ and}$$

$$g(x_{n,i+1}) = g(x_{ni}) + g'(x_{ni})(1/n) + g''(x_{ni}^*)(1/n)^2/2!,$$

where x_{ni}^o is in $[x_{n,i-1}, x_{ni}]$ and x_{ni}^* is in $[x_{ni}, x_{n,i+1}]$. Since the

first derivative terms cancel out, we obtain

$$g(x_{n,i-1}) + g(x_{n,i+1}) = 2g(x_{ni}) + (1/2n^2)[g''(x_{ni}^o) + g''(x_{ni}^*)]. \quad (2.4.14)$$

Substitution into (2.4.12), factoring out $(1/n^2)$, and rearranging

yields

$$n e^{a/n} S_n = (1/n) \Sigma_{i=2}^{n-1} g(x_{ni})\{(a^2 + O(n^{-2}))g(x_{ni}) - (1/2)[g''(x_{ni}^o) + g''(x_{ni}^*)]\} \quad (2.4.15)$$

The last quantity converges to

$$\int_0^1 g(x)[a^2 g(x) - g''(x)]dx \quad (2.4.16)$$

by the continuity of $g''(x)$ and the fact that (2.4.15) has the form of

an approximating Riemann sum. Finally, the result for S_n follows from

the fact that

$$[n\, e^{a/n}]^{-1}(1-\rho_n^2)^{-1} = [(1/n)e^{-a/n}]/[1-\exp(-2a/n)] \rightarrow 1/2a$$

by an application of L'Hospital's Rule.

Next consider the boundary term B_{n1}, which by an argument

similar to that above may be expressed as

$$B_{n1} = g(x_{n1})[g(x_{n1}) - \rho_n g(x_{n2})]$$

$$= (1/n)\, g(x_{n1})[-\, g'(x_{n1}^*) + a\, g(x_{n2}) + g(x_{n2})\, O(1/n)], \quad (2.4.17)$$

where x_{n1}^* is in $[x_{n1}, x_{n2}]$. Since $x_{ni} \rightarrow 0$ for i fixed (as $n \rightarrow \infty$) and

g and its first derivative are continuous, we have

$$g(x_{n1}) \rightarrow g(0), \; g(x_{n2}) \rightarrow g(0), \text{ and } g'(x_{n1}^*) \rightarrow g'(0+), \text{ as } n \rightarrow \infty.$$

The result for B_{n1} follows by noting that

$$(1/n)(1-\rho_n^2)^{-1} \rightarrow (1/2a),$$

from another application of L'Hospital's Rule. A similar argument

produces the result for B_{nn}, and is hence omitted. ∎

One should note the similarity of this result with the inner product

representation of a Hilbert space investigated in Parzen (1967,p.403)

in the context of regression analysis of a continuous parameter time

series. Now we are equipped to easily prove the main result of this

section.

Theorem (2.4.1) Suppose m is fixed, g is not identically zero, and g

satisfies the conditions in Lemma (2.4.1). Then for the Ornstein-

Uhlenbeck model (2.1.7), augmented to include $\rho = 0$, a necessary and

sufficient condition for $M(W_n^*) \rightarrow 0$ as $n \rightarrow \infty$ is $\rho = 0$ ($a = \infty$).

Proof: Suppose that $\rho = 0$. Then $\Gamma = I$ and has an eigenvalue 1 of

multiplicity n. Hence $n/e_{n1}(\Gamma) = n \rightarrow \infty$ as $n \rightarrow \infty$ and $M(W_n^*) \rightarrow 0$ by

Theorem (2.3.1). [Alternatively, an examination of (2.3.1) shows that

$M(W_n^*) \sim \sigma^2/(mn) \rightarrow 0$ as $n \rightarrow \infty$.]

Next assume $M(W_n^*) \rightarrow 0$. For the sake of a contradiction, assume

$\rho > 0$. Define the quantity $J(a)$ to be

$$J(a) = (1/2a)\{\int_0^1 g(x)[a^2g(x)-g''(x)]dx$$

$$+ g(0)[ag(0)-g'(0+)] + g(1)[ag(1)+g'(1-)]\}. \qquad (2.4.18)$$

By Lemma (2.4.1) we have $g'\Gamma^{-1}g \rightarrow J(a) < \infty$. Recall from (2.3.1) that

$$M(W_n^*) = [(g'g)/n] \, [1 + (m/\sigma^2)g'\Gamma^{-1}g]^{-1} \rightarrow 0,$$

thereby requiring that $g'\Gamma^{-1}g \rightarrow \infty$ (since $[(g'g)/n]$ approaches a finite, positive constant). We therefore have a contradiction and the theorem is proved. ∎

2.5 Some Aspects of Spectral Theory

In this section we will digress briefly to examine how the spectral theory corresponding to the Ornstein-Uhlenbeck covariance structure relates to the problem at hand. The interplay of the spectral density and the eigenvalues of the correlation matrix may help us better understand the conditions of Theorem (2.3.1). We will examine the spectral density, $f(t)$, of a stationary process having the covariance structure (2.1.6). In the equally-spaced case, this coincides with an autoregressive process of order 1, that is, an AR(1) process. Recall that the (normalized) spectral density, $f(t)$, is the Fourier transform of

$$\gamma(u) \equiv \text{corr}[\bar{y}(x),\bar{y}(x+u)] = \exp(-a|u|), \qquad (2.5.1)$$

namely,

$$f(t) \equiv (1/2\pi) \int_{-\infty}^{\infty} \exp(-iut) \, \gamma(u)du. \qquad (2.5.2)$$

The autocorrelation function may in turn be recovered through the relation,

$$\gamma(u) = \int_{-\infty}^{\infty} \exp(iut)\, f(t)dt. \qquad (2.5.3)$$

Since γ is even, f is real-valued and we may write (2.5.2) as

$$f(t) = (1/2\pi) \int_{-\infty}^{\infty} \cos(tu)\, \gamma(u)du$$

$$= (1/\pi) \int_{0}^{\infty} \exp(-au)\cos(tu)du$$

$$= (1/\pi)\, \mathrm{Re}\{\int_{0}^{\infty} \exp[-u(a-it)]du\}$$

$$= (1/\pi)\, \mathrm{Re}\{1/(a-it)\}$$

$$= a/\pi(a^2 + t^2). \qquad (2.5.4)$$

Note that f(t) is symmetric with

$$\max_{\{t\}} f(t) = f(0) = 1/\pi a. \qquad (2.5.5)$$

Now for discrete-time processes, (2.5.2) may alternatively be written

$$f(t) = (1/2\pi) \Sigma_{v=-\infty}^{\infty} \gamma(v) \exp(-ivt), \qquad (2.5.6)$$

$$= (1/2\pi)\, [1 + 2\, \Sigma_{v=1}^{\infty} \gamma(v) \cos(vt)], \quad t \text{ in } [-\pi,\pi]. \qquad (2.5.7)$$

In our development we observe that the unit spacing $1/n$ tends to 0 as $n \to \infty$. Replacing ρ with ρ_n and a with a/n, we obtain the following analogs:

$$\gamma_n(v) = \exp(-a|v|/n) = \rho_n^{|v|}, \quad v=0,\pm 1,\pm 2,\ldots \qquad (2.5.8)$$

$$f_n(t) = (1/2\pi)[1 + 2\, \Sigma_{v=1}^{\infty} \gamma_n(v) \cos(vt)]$$

$$= (1/2\pi)(1-\rho_n^2)\, |1 - \rho_n e^{-it}|^{-2} \qquad (2.5.9)$$

$$\leq (1/2\pi)(1-\rho_n^2)(1-\rho_n)^{-2}$$

$$= (1/2\pi)[(1+\rho_n)/(1-\rho_n)]$$

$$= (1/2\pi)\, M_n(\rho), \text{ say.} \qquad (2.5.10)$$

Note that $M_n(\rho)$ becomes unbounded as $n \to \infty$ unless $\rho = 0$, in which case it is identically 1. We will use the above results in subsequent sections, but first we establish some technical details in the form of a lemma.

<u>Lemma (2.5.1)</u> -- Grenander and Szego (1958)

Let f be an integrable function on $[-\pi,\pi]$ satisfying

$$m \leq f(t) \leq M, \text{ for all t in } [-\pi,\pi]. \qquad (2.5.11)$$

Let

$$\gamma(u) = \int_{-\pi}^{\pi} \exp(iut) \, f(t)dt \qquad (2.5.12)$$

and

$$\Gamma = \text{TOEPL}[\gamma(1-n),\ldots,\gamma(0),\ldots,\gamma(n-1)]. \qquad (2.5.13)$$

Then there exists a constant c for which

$$c \, m \leq e_{nn}(\Gamma) \leq e_{n1}(\Gamma) \leq c \, M.$$

For g satisfying the conditions of Theorem (2.4.1) in the Ornstein-Uhlenbeck model, we know that $n/e_{n1}(\Gamma) \not\longrightarrow \infty$, else Theorem (2.3.1) would be incorrect. Computer studies suggest that $n/e_{n1}(\Gamma)$ approaches a finite limit, which undoubtedly is a relatively simple function of ρ, which the authors were unable to discover. However, we know that

$$n/e_{n1}(\Gamma) \geq n/[c \, M_n(\rho)] \longrightarrow -(1/2c)\log(\rho) = a/2c, \qquad (2.5.14)$$

which provides a lower bound for this limit.

The reader may question whether there exists any covariance function γ, independent of n, which would satisfy the eigenvalue condition of Theorem (2.3.1), other than the uncorrelated case. We conjecture there are none, but this assertion would need investigation. Of course, if γ is allowed to change with n, then the condition can clearly be met. For instance, if $\gamma_n(u)$ (for $u \neq 0$) tends to zero sufficiently fast, then one would expect the condition to be met. On the other hand, if one recalls the natural application

of growth curves, it seems that for fixed m, a larger value of n
would provide more accurate information about the m-sample of
animals, but not about the entire population, of which $g(x)$
represents the true average response at time x. Hence for reasonable
models and fixed m, the absence of a consistent estimator seems
evident.

2.6 Pointwise Expressions at the True Optimum

Recall that our estimator has the form

$$g_n(x) = \Sigma_i \ w_{ni}(x) \ \bar{y}(x_i). \qquad (2.6.1)$$

We may obtain the asymptotic behavior of this estimator at a point
arbitra. ly close to x by choosing $j = j_n$ such that

$$x_{nj} = (j-1)/n \longrightarrow x, \text{ as } n \longrightarrow \infty,$$

which, in formula (2.6.1), becomes

$$g_n(x_{nj}) = \Sigma_i \ w_{ni}(x_{nj}) \ \bar{y}(x_{ni}). \qquad (2.6.2)$$

Of course, we could not hope to do better in a global sense than if
we knew the true correlation and the true function and used

$$g_n^{\bullet}(x_{nj}) = \Sigma_i \ w_{ni}^{\bullet}(x_{nj}) \ \bar{y}(x_{ni}) \qquad (2.6.3)$$

where $w_{ni}^{\bullet}(x_{nj})$ is the $(i,j)^{th}$ element of the MASE-minimizing matrix

$$W_n^{\bullet} = (\Sigma_m + gg')^{-1} gg' = c_{mn} \ \Gamma^{-1} gg', \qquad (2.6.4)$$

and the scalar quantity c_{mn} is defined by

$$c_{mn} = [(\sigma^2/m) + g'\Gamma^{-1}g]^{-1}. \qquad (2.6.5)$$

We have shown that the optimum global measure of error, $M(W_n^{\bullet})$, does
not approach zero and hence (2.6.3) fails to be globally consistent

under mild restrictions on g in the Ornstein-Uhlenbeck model. Of course we note that (2.6.3) is not an estimator since the weights depend on the very function we are attempting to estimate. We will abuse the terminology and refer to this quantity as an estimator when it is actually the optimal linear combination of the subject means. Note that the jth column of $W_n^{\bullet}$ is the vector of weights used in (2.6.3) when "predicting" at x_{nj}. From (2.6.4) we see that this is the vector

$$w_n^{\bullet}(x_{nj}) = [w_{n1}^{\bullet}(x_{nj}), \ldots, w_{nn}^{\bullet}(x_{nj})]' \qquad (2.6.6)$$

$$= g(x_{nj}) \, c_{mn} \, \Gamma^{-1}g. \qquad (2.6.7)$$

Now $\Gamma^{-1}g$ is recoverable from equations (2.4.6), (2.4.7), and (2.4.8) as given in Lemma (2.4.1), yielding a formula for the $(i,j)^{th}$ element of $W_n^{\bullet}$:

$$w_{ni}^{\bullet}(x_{nj}) = g(x_{nj}) \, c_{mn} \, (1-\rho_n^2)^{-1} \, d_{ni}, \qquad (2.6.8)$$

where the quantity d_{ni} is

$$d_{ni} = \{(1+\rho_n^2)g(x_{ni}) - \rho_n[g(x_{n,i-1})+g(x_{n,i+1})]\} \qquad (2.6.9)$$

for $i=2,3,\ldots,n-1$, with the end values being

$$d_{n1} = \{g(x_{n1}) - \rho_n g(x_{n2})\}, \quad d_{nn} = \{g(x_{nn}) - \rho_n g(x_{n,n-1})\}. (2.6.10)$$

We next evaluate certain pointwise expressions, incorporating the true optimum global values. For reference, we will state the result in the following lemma.

Lemma (2.6.1) Suppose g, σ^2, and Γ are known, and, we use (2.6.3) to estimate $g(x_{nj})$. Then for c_{mn} defined by (2.6.5), we have

(a) $E[g_n^*(x_{nj})] = g(x_{nj}) \, c_{mn} \, g'\Gamma^{-1}g$ (2.6.11)

(b) $\text{Bias}[g_n^*(x_{nj}), g(x_{nj})] = -\, g(x_{nj}) \, (\sigma^2/m) \, c_{mn}$ (2.6.12)

(c) $\text{Var}[g_n^*(x_{nj})] = g^2(x_{nj}) \, (\sigma^2/m) \, c_{mn}^2 \, g'\Gamma^{-1}g$ (2.6.13)

(d) $\text{MSE}[g_n^*(x_{nj})] = g^2(x_{nj}) \, (\sigma^2/m) \, c_{mn} = g^2(x_{nj}) \, [1+g'\Sigma_m^{-1}g]^{-1}$ (2.6.14)

Proof: For (a), we use (2.6.7) to obtain

$$E[g_n^*(x_{nj})] = \Sigma_i \, w_{ni}^*(x_{nj}) \, g(x_{ni})$$

$$= [g(x_{nj}) \, c_{mn} \, \Gamma^{-1}g]' \, g = g(x_{nj}) \, c_{mn} \, g'\Gamma^{-1}g.$$

For (c), we have

$$\text{Var}[g_n^*(x_{nj})] = [g(x_{nj}) \, c_{mn} \, \Gamma^{-1}g]' \, \text{Var}(\bar{y}) \, [g(x_{nj}) \, c_{mn} \, \Gamma^{-1}g],$$

from which (c) follows since $\text{Var}(\bar{y}) = (\sigma^2/m)\Gamma$. Properties (b) and (d) follow easily from (a) and (c) since

$$\text{Bias}[g_n^*(x_{nj}), g(x_{nj})] = E[g_n^*(x_{nj})] - g(x_{nj}), \text{ and}$$

$$\text{MSE}[g_n^*(x_{nj})] = \text{Var}[g_n^*(x_{nj})] + \text{Bias}^2[g_n^*(x_{nj}), g(x_{nj})]. \quad \blacksquare$$

Next we obtain the limiting form of these quantities under certain conditions on the regression function g.

Theorem (2.6.1) Suppose g satisfies the conditions of Lemma (2.4.1) and g is not identically 0. Let $J(a)$ denote the limit as $n \to \infty$ of $g'\Gamma^{-1}g$, as given in the lemma (with $\rho \neq 0$), that is,

$$J(a) = (1/2a)\{\int_0^1 g(x)[a^2g(x)-g''(x)]dx$$

$$+ g(0)[ag(0)-g'(0+)] + g(1)[ag(1)+g'(1-)]\}, \quad (2.6.15)$$

which is finite since $a < \infty$. Let $J_m(a)$ denote the limit of $g'\Sigma_m^{-1}g$, as $n \rightarrow \infty$, which is

$$J_m(a) = (m/\sigma^2)J(a), \qquad (2.6.16)$$

and suppose

$$x_{nj} = (j-1)/n \rightarrow x, \text{ as } n \rightarrow \infty. \qquad (2.6.17)$$

Then we have, as $n \rightarrow \infty$,

(a) $E[g_n^*(x_{nj})] \rightarrow g(x) J_m(a)/(1 + J_m(a))$, $\qquad (2.6.18)$

(b) $\text{Bias}[g_n^*(x_{nj}),g(x_{nj})] \rightarrow - g(x)/(1 + J_m(a))$, $\qquad (2.6.19)$

(c) $\text{Var}[g_n^*(x_{nj})] \rightarrow g^2(x) J_m(a)/(1 + J_m(a))^2$, $\qquad (2.6.20)$

(d) $\text{MSE}[g_n^*(x_{nj})] \rightarrow g^2(x)/(1 + J_m(a))$. $\qquad (2.6.21)$

<u>Proof</u>: The algebra follows easily from Lemma (2.6.1). by noting

$$(\sigma^2/m) \, c_{mn} = (\sigma^2/m)[(\sigma^2/m) + g'\Gamma^{-1}g]^{-1} = [1+g'\Sigma_m^{-1}g]^{-1}$$

$$\rightarrow (\sigma^2/m)[(\sigma^2/m) + J(a)]^{-1} = (1 + J_m(a))^{-1}. \qquad \blacksquare$$

In view of this result and the fact that for fixed $a > 0$,

$$J_m(a) \rightarrow \infty \text{ if and only if } m \rightarrow \infty, \qquad (2.6.22)$$

and for fixed m,

$$J_m(a) \rightarrow \infty \text{ if and only if } a \rightarrow \infty \ (\rho = 0),$$

it is clear that the pointwise estimator evaluated at the optimum global weights will fail to be mean-square consistent (under reasonable conditions on g) unless $m \rightarrow \infty$, or we have an uncorrelated-errors model. Indeed, for a fixed value of m, this estimator is not even asymptotically unbiased when correlation exists in the model. Note also that for $g(x) \neq 0$ and $n \rightarrow \infty$,

$$\text{Bias}^2[g_n^*(x_{nj}),g(x_{nj})]/\text{Var}[g_n^*(x_{nj})] \rightarrow 1/J_m(a), \qquad (2.6.23)$$

which in turn approaches zero if we let $m \rightarrow \infty$ or $a \rightarrow \infty$. Hence the contribution of the bias to the mean squared error at a point becomes increasingly small as the sample size from the population of subjects increases. This is consistent with our observation in the proof of Theorem (2.4.1) that, for $\rho = 0$, $M(W_n^*) \sim \sigma^2/mn = (1/n)\text{Var}[\bar{y}(x_j)]$. We note in passing that $\text{MSE}[g_n^*(x_{nj})]$ must be the minimum MSE at the point x_{nj} among all linear estimates, else Theorem (2.2.1) is contradicted.

2.7 Approaches Using Continuous Realizations

In terms of averages, the model (2.1.1) may be written in a condensed form,

$$\bar{y}(x_t) = g(x_t) + \bar{e}(x_t), \quad t=1,2,\ldots,n,$$

$$E[\bar{e}(x_t)] = 0,$$

$$\text{cov}[\bar{e}(x_t),\bar{e}(x_v)] = R(x_t,x_v) = (\sigma^2/m) \Gamma(x_t,x_v), \quad (2.7.1)$$

where we will take the $(t,v)^{th}$ element of the correlation matrix Γ to be

$$\Gamma(x_t,x_v) = \gamma(x_t-x_v), \quad (2.7.2)$$

which in the Ornstein-Uhlenbeck model is

$$\Gamma(x_t,x_v) = \exp\{-a|x_t-x_v|\}. \quad (2.7.3)$$

Define continuous data realizations (or records) over the interval $[0,1]$ to be the random functions

$$\{\bar{y}(x): \text{ all } x \text{ in } [0,1]\}, \quad (2.7.4)$$

$$\{\bar{e}(x): \text{ all } x \text{ in } [0,1]\}. \quad (2.7.5)$$

Our estimator may be written

$$g_n(x) = \Sigma_i \ w_{ni}(x) \ \bar{y}(x_{ni})$$

$$= (1/n) \ \Sigma_i \ [nw_{ni}(x)] \ \bar{y}(x_{ni}). \tag{2.7.6}$$

Suppose that the weights $nw_{ni}(x)$ are of the form

$$nw_{ni}(x) = K_n(x,x_{ni}). \tag{2.7.7}$$

Observe that kernel estimators may be written explicitly in this form. We may write

$$g_n(x) = (1/n) \ \Sigma_i \ K_n(x,x_{ni}) \ \bar{y}(x_{ni})$$

$$= \int_0^1 K_n(x,s) \ \bar{y}(s) \ dF_n(s). \tag{2.7.8}$$

If g is absolutely integrable, $E|\bar{e}(x_t)| < \infty$ for all t, and $K_n(x,\cdot)$ converges uniformly to $K(x,\cdot)$, it may be shown that (2.7.8) may be written as

$$\int_0^1 K(x,s) \ \bar{y}(s) \ dF_n(s) + o_p(1). \tag{2.7.9}$$

Furthermore, if $K(x,\cdot)\bar{y}(\cdot)$ is continuous (wp1), the integral in (2.7.9) approaches

$$\int_0^1 K(x,s) \ \bar{y}(s) \ f(s) \ ds, \ \text{as} \ n \to \infty. \tag{2.7.10}$$

Recall that $f(s)$ is the design density, which in our situation is identically one for the uniformly spaced design. Hence for a continuous data record, ideally with $n = \infty$ data points available, one might be led to consider an estimator of the form

$$g_\infty(x) = \int_0^1 K(x,s) \ \bar{y}(s) ds. \tag{2.7.11}$$

We find the mean, variance and mean squared error of this estimator in the following lemma.

Lemma (2.7.1) Suppose we consider the estimator (2.7.11) in the model
(2.7.1) where

$$E[\bar{e}(x)] = 0 \text{ and } cov[\bar{e}(x),\bar{e}(s)] = R(x,s). \qquad (2.7.12)$$

Then

(i) $E[g_e(x)] = \int K(x,s)g(s)ds,$ \qquad (2.7.13)

(ii) $Var[g_e(x)] = \iint K(x,s)K(x,t)R(s,t) \, ds \, dt$ \qquad (2.7.14)

(iii) $MSE[g_e(x)] = \iint K(x,s)K(x,t)R(s,t) \, ds \, dt$

$$+ \ [\int K(x,s)g(s)ds - g(x)]^2. \qquad (2.7.15)$$

Proof: Clearly (i) follows from Fubini's Theorem. For (ii) consider

$$E[g_e^2(x)] = E[\int K(x,s)\bar{y}(s)ds \int K(x,t)\bar{y}(t)dt]$$

$$= \iint K(x,s)K(x,t)E[\bar{y}(s)\bar{y}(t)] \, ds \, dt$$

$$= \iint K(x,s)K(x,t)[R(s,t) + g(s)g(t)] \, ds \, dt.$$

Subtracting $E^2[g_e(x)]$ from the last expression yields (ii). The
expression (iii) follows directly since it equals the sum of the
variance and the squared bias. ∎

One would like to find the function $K(x,\cdot)$ which minimizes

$$MSE[g_e(x)] = E[g_e(x) - g(x)]^2. \qquad (2.7.16)$$

Using an approach based on the calculus of variations, one may prove
the following theorem.

Theorem (2.7.1) Consider the model (2.7.1) and the continuous data record estimator (2.7.11). Let ϕ be defined implicitly via the integral equation

$$g(x) = \int_0^1 R(x,u)\, \phi(u)\, du, \qquad (2.7.17)$$

and define the quantities

$$L = \int_0^1 g(u)\, \phi(u)\, du, \qquad (2.7.18)$$

$$K^*(x,s) = \text{the minimizer of } MSE[g_\infty(x)], \qquad (2.7.19)$$

$$g_\infty^*(x) = \int_0^1 K^*(x,s)\, \bar{y}(s)\, ds. \qquad (2.7.20)$$

Then

(i) $K^*(x,s) = g(x)\phi(s)/(1 + L)$, $\qquad (2.7.21)$

(ii) $g_\infty^*(x) = g(x)\,[\,\int \phi(s)\bar{y}(s)ds\,]/(1 + L)$, $\qquad (2.7.22)$

(iii) $E[g_\infty^*(x)] = g(x)\, L/(1 + L)$, $\qquad (2.7.23)$

(iv) $Bias[g_\infty^*(x), g(x)] = -\,g(x)/(1 + L)$ $\qquad (2.7.24)$

(v) $Var[g_\infty^*(x)] = g^2(x)\, L/(1 + L)^2$ $\qquad (2.7.25)$

(vi) $MSE[g_\infty^*(x)] = g^2(x)/(1 + L)$. $\qquad (2.7.26)$

Proof: The proof of this result was obtained from T. E. Wehrly, but is here omitted. ∎

The reader should note the analogy between the general results of this theorem and those of Theorem (2.6.1), which deals with the special case of the Ornstein-Uhlenbeck model. Observe that the previous theorem dealt with the limit of the global optimum for a finite number of points n, as $n \to \infty$, whereas the present theorem first allows $n \to \infty$ and then obtains the optimum at a given point x.

An interesting question is whether $J_m(a)$ of Theorem (2.6.1) is the same as L of the present theorem under the Ornstein-Uhlenbeck model. If so, we observe that the evaluation of finite-n estimators at global optimizers and letting $n \rightarrow \infty$ leads to the same results as that of optimizing a continuous record estimator at a point. This should be true since the form of the estimator is global in the sense that it weights the observations across the entire interval.

It may be possible to directly extract ϕ in equation (2.7.15) by the theory of integral equations, as outlined in Whittaker and Watson (1963,ch. 11). Since the nucleus of the equation, $R(x,u)$ is symmetric, that is,

$$R(x,u) = R(u,x) = (\sigma^2/m) \exp\{-a|x-u|\}, \qquad (2.7.27)$$

a solution for ϕ is given by

$$\phi(x) = \Sigma_{n=1}^{\infty} a_n \lambda_n \phi_n(x), \qquad (2.7.28)$$

provided this series converges uniformly. The quantities λ_n, $\phi_n(x)$ are the so-called characteristic numbers and functions of the nucleus $R(x,u)$. These quantities are also known as eigenvalues and eigenfunctions, respectively. The quantity $\Sigma \, a_n \phi_n(x)$ represents the expansion of $g(x)$. The reader is referred to Whittaker and Watson (1963,p. 231) for more details.

Mere calculation of the characteristic numbers may be quite a prodigious task. The characteristic numbers are the roots of the equation

$$D(\lambda) = 1 + \Sigma_{n=1}^{\infty} a_n \lambda^n/n!, \qquad (2.7.29)$$

where

$$a_n = (-1)^n \int_a^b \int_a^b \cdots \int_a^b \det(R) \, du_1, \ldots, du_n, \qquad (2.7.30)$$

and the $(i,j)^{th}$ element of the (n by n) matrix R (in the Ornstein-Uhlenbeck model) is

$$R(u_i, u_j) = (\sigma^2/m) \exp\{-a|u_i - u_j|\}. \qquad (2.7.31)$$

Fortunately, there exist other means of finding a solution for ϕ, provided g satisfies certain regularity conditions. One solution, based on Fourier analysis, may require fairly stringent conditions on g which essentially "assume away" the difficulty posed by the boundary terms which are present in Theorem (2.4.1). We will not be concerned about this difficulty since Theorem (2.4.1) already takes these into account. The main emphasis will be to show that identical results are obtained, although in a more restricted case.

Theorem (2.7.2) Suppose

$$g(x) = \int_{-\infty}^{\infty} R(x, s) \, \phi(s) \, ds, \qquad (2.7.32)$$

where R is of the form R(x-s) and

$$R(u) = (\sigma^2/m) \exp(-a|u|). \qquad (2.7.33)$$

Assume also that g vanishes outside [0,1] and is everywhere twice continuously differentiable. We then have

$$\phi(s) = (m/2a\sigma^2)[a^2 g(s) - g''(s)], \qquad (2.7.34)$$

and hence the optimum kernel is

$$K^*(x, s) = g(x)[a^2 g(s) - g''(s)] / [(2a\sigma^2/m) + \int_0^1 g(u)[a^2 g(u) - g''(u)] du]. \qquad (2.7.35)$$

Proof: We have the convolution

$$g(x) = R*\phi(x) = \int_{-\infty}^{\infty} R(x-s) \, \phi(s) \, ds. \qquad (2.7.36)$$

Let $f^{\dagger}$ be the Fourier transform of the function f. Then by a well known property of convolutions,

$$g^{\dagger} = (R*\phi)^{\dagger} = R^{\dagger} \, \phi^{\dagger}. \qquad (2.7.37)$$

It is also well known that for R of the form (2.7.33),

$$R^{\dagger}(u) = (2a\sigma^2/m)(a^2 + u^2)^{-1}. \qquad (2.7.38)$$

We now have

$$\phi(s) = (1/2\pi) \int_{-\infty}^{\infty} e^{-isu}[g^{\dagger}(u)/R^{\dagger}(u)]du$$

$$= (m/2a\sigma^2)[(1/2\pi)a^2 \int e^{-isu}g^{\dagger}(u)du + (1/2\pi) \int u^2 e^{-isu}g^{\dagger}(u)du]$$

$$= (m/2a\sigma^2)[a^2 g(s) - g''(s)], \qquad (2.7.39)$$

where the last equality results from differentiating the inverse Fourier transform of $g^{\dagger}$ twice under the integral sign. Since g vanishes outside $[0,1]$, so does ϕ, and it follows that

$$g(x) = \int_0^1 R(x-s) \, \phi(s) \, ds.$$

The proof is completed by applying Theorem (2.7.1). ∎

Define the quantity, $J_m^*(a)$ to be the expression (2.6.16) for $J(a)$ without the boundary terms:

$$J_m^*(a) = (m/2a\sigma^2) \int_0^1 g(s)[a^2 g(s)-g''(s)]ds. \qquad (2.7.40)$$

Observe that the boundary terms

$$(m/2a\sigma^2) \, g(0)[ag(0)-g'(0+)], \qquad (2.7.41)$$

$$(m/2a\sigma^2) \, g(1)[ag(1)+g'(1-)], \qquad (2.7.42)$$

vanish if $g(0) = g(1) = 0$, which is a condition of Theorem (2.7.2). Also note that $K^*(x,s)$ in Theorem (2.7.2) may be expressed as

$$K^*(x,s) = g(x)\phi(s)/[1 + J_m^*(a)].$$

From this it is easy to see that the results of Theorems (2.7.1) and
(2.7.2) are the same, with $J_m(a)$ replaced by $J_m^*(a)$, provided g
satisfies certain boundary conditions.

2.8   Sufficient Conditions for Mean-Square Consistency

The purpose of this section is to prove a theorem which gives
sufficient conditions for which

$$E[g_n(x) - g(x)]^2 \to 0, \text{ as } n \to \infty \qquad (2.8.1)$$

at a point x. Recall that estimators which satisfy this property are
said to be mean-square consistent estimators. It is sufficient to
demonstrate that the variance and the squared bias tend to zero as
$n \to \infty$. Earlier we discovered that if the Ornstein-Uhlenbeck
correlation model holds, the estimator g_n^* is not consistent for g as
$n \to \infty$. Obviously the Ornstein-Uhlenbeck function will violate the
conditions of this theorem. Indeed, for a correlation function not
dependent on n, it is not clear at this point whether or not the
theorem will admit any consistent estimators at all when m is fixed.
However, it will later be shown that the theorem will admit
estimators in the uncorrelated case. If one were willing to assume
that the correlation function changes with n, then consistency is
clearly possible. The approach outlined below is given by Georgiev
and Greblicki (1986) in the case of uncorrelated-errors and m=1.

<u>Theorem (2.8.1)</u> Suppose the model (2.1.1) holds and fix m. Consider estimators of the form (2.1.8), where the weights may depend on the entire design vector. Let $C[0,1]$ denote the space of uniformly continuous functions defined on the interval $[0,1]$. Suppose for a given x in $[0,1]$,

(i) $\sup_n \Sigma_i \, |w_{ni}(x)| \leq B < \infty,$ (2.8.2)

(ii) $\Sigma_i \, w_{ni}(x) \to 1$, as $n \to \infty,$ (2.8.3)

(iii) for all $\epsilon > 0$, $\Sigma_i \, |w_{ni}(x)| \, I\{|x-x_{ni}| > \epsilon\} \to 0$, as $n \to \infty$ (2.8.4)

where $I\{\cdot\}$ is an indicator function, and

(iv) $w_n'(x) \, \Gamma \, w_n(x) = \Sigma_i \Sigma_j \, w_{ni}(x) w_{nj}(x) \gamma(x_i - x_j) \to 0,$

as $n \to \infty$, where $w_n(x) = [w_{n1}(x), \ldots, w_{nn}(x)]'.$ (2.8.5)

Then for $g \in C[0,1]$, $E[g_n(x) - g(x)]^2 \to 0$, as $n \to \infty$.

<u>Proof</u>: Let $\epsilon > 0$ and x be fixed.

For the bias consider $|Eg_n(x) - g(x)|$

$= |\Sigma_i w_{ni}(x) g(x_{ni}) - g(x)|$

$= |\Sigma_i w_{ni}(x) [g(x_{ni}) - g(x) + g(x)] - g(x)|$

$= |\Sigma_i w_{ni}(x) [g(x_{ni}) - g(x)] + g(x) [\Sigma_i w_{ni}(x) - 1]|$

$\leq \Sigma_i |w_{ni}(x)| |g(x_{ni}) - g(x)| + |g(x)| |\Sigma_i w_{ni}(x) - 1|$

$= \Sigma_i |w_{ni}(x)| |g(x_{ni}) - g(x)| I\{|x-x_{ni}| \leq \epsilon\}$

$\qquad + \Sigma_i |w_{ni}(x)| |g(x_{ni}) - g(x)| I\{|x-x_{ni}| > \epsilon\}$

$\qquad\qquad + |g(x)| |\Sigma_i w_{ni}(x) - 1|$

$\leq \Sigma_i |w_{ni}(x)| \sup_{[0,1]} [\, |g(x_{ni}) - g(x)| I\{|x-x_{ni}| \leq \epsilon\} \,]$

$\qquad + \Sigma_i [|g(x_{ni})| + |g(x)|] \, |w_{ni}(x)| \, I\{|x-x_{ni}| > \epsilon\}$

$\qquad\qquad + \sup_z |g(z)| \, |\Sigma_i w_{ni}(x) - 1|$

$$\leq \sup_{|x-z|\leq \epsilon} \; |g(x)-g(z)| \; \Sigma_i |w_{ni}(x)|$$

$$+ \; 2 \sup_z |g(z)| \; \Sigma_i |w_{ni}(x)| I\{|x-x_{ni}|>\epsilon\}$$

$$+ \; \sup_z |g(z)| \; |\Sigma_i w_{ni}(x) - 1|$$

$$\leq B \sup_{|x-z|\leq \epsilon} \; |g(x)-g(z)|$$

$$+ \; 2 \sup_z |g(z)| \; \Sigma_i |w_{ni}(x)| I\{|x-x_{ni}|>\epsilon\}$$

$$+ \; \sup_z |g(z)| \; |\Sigma_i w_{ni}(x) - 1|. \qquad (2.8.6)$$

Now the second and third terms may be made as small as desired by choosing n sufficiently large while the first term may be made arbitrarily small by letting $\epsilon \to 0$ (by the uniform continuity of g). Hence the absolute bias,

$$|Eg_n(x) - g(x)| \to 0, \text{ as } n \to \infty.$$

Therefore the squared bias must also approach zero. The result follows by noting

$$Var[g_n(x)] = (\sigma^2/m) \; w_n'(x) \; \Gamma \; w_n(x) \to 0, \text{ as } n \to \infty, \qquad (2.8.7)$$

using assumption (iv). ∎

Observe that assumption (iii) requires that the sum of the contribution of weights at points x_{ni} (when predicting at x) die off for x_{ni}-values an arbitrary distance from x (as n gets large). This says that the estimation at point x may be concentrated on the weights attached to y-values corresponding to x_{ni}-values in smaller and smaller neighborhoods of x as n gets larger. We will later see that this is what happens in the special case of the kernel form $w_{ni}(x;h)$ as the bandwidth parameter h tends to 0 for increasing n. The following corollary simplifies some of the conditions of this

theorem at the expense of being more restrictive.

Corollary (2.8.1) Consider the conditions of Theorem (2.8.1) and let $e_{n1}(\Gamma)$ denote the largest eigenvalue of Γ. Then

(a) Condition (i) is implied by Condition (ii) for nonnegative weights, $w_{ni}(x)$, for all n, i, and x.

(b) Conditions which imply Condition (iv) are

$$e_{n1}(\Gamma) \Sigma_i \, w_{ni}^2(x) \rightarrow 0, \text{ as } n \rightarrow \infty, \qquad (2.8.8)$$

or the even stronger condition,

$$n \, \Sigma_i \, w_{ni}^2(x) \rightarrow 0, \text{ as } n \rightarrow \infty. \qquad (2.8.9)$$

(c) If $e_{n1}(\Gamma)$ is bounded above for all n (as in the uncorrelated case), Condition (iv) may be replaced by the weaker condition

$$\Sigma_i \, w_{ni}^2(x) \rightarrow 0, \text{ as } n \rightarrow \infty. \qquad (2.8.10)$$

(d) For $m \rightarrow \infty$, Condition (iv) is implied by

$$(n/m) \, \Sigma_i \, w_{ni}^2(x) \rightarrow 0, \text{ as } n \rightarrow \infty. \qquad (2.8.11)$$

Proof: For (a), note that

$$\Sigma_i \ |w_{ni}(x)| = \Sigma_i \ w_{ni}(x),$$

which must be bounded in order for the limit to exist. For (b), recall

$$(\sigma^2/m) \ w_n'(x) \ \Gamma \ w_n(x) \leq (\sigma^2/m) \ e_{n1}(\Gamma) \ w_n'(x)w_n(x) \qquad (2.8.12)$$

$$\leq (\sigma^2/m) \ n \ w_n'(x)w_n(x), \qquad (2.8.13)$$

since the sum of the (nonnegative) eigenvalues of Γ equals the trace of Γ, which has value 1 along the diagonal. We are done since (c) and (d) follow directly from (b). ∎

Corollary (2.8.2) Consider Theorem (2.8.1). For the correlation model (2.1.7), Condition (iv) is implied by

$$M_n(\rho) \ \Sigma_i \ w_{ni}^2(x) \rightarrow 0, \text{ as } n \rightarrow \infty. \qquad (2.8.14)$$

Hence for $\rho = 0$, this is equivalent to

$$\Sigma_i \ w_{ni}^2(x) \rightarrow 0, \text{ as } n \rightarrow \infty.$$

Proof: The result follows from the Corollary (2.8.1) by recalling

$$M_n(\rho) = (1 + \rho^{1/n})/(1 - \rho^{1/n}) \qquad (2.8.15)$$

is an upper bound for $e_{n1}(\Gamma)$, apart from a multiplicative constant free of n, and equals one for $\rho = 0$. ∎

We will now proceed to Chapter 3, where weights of the kernel type are considered. The results of this chapter will be summarized in Section 3.1 to motivate the approach taken for the numerical study.

3. KERNEL ESTIMATORS OF REGRESSION FUNCTIONS

3.1 Introduction

A summary of the main points of Chapter 2 will lead us to the goal of Chapter 3 and serve as a useful reference. Throughout this chapter the term <u>correlated errors model</u> will refer to model (2.1.1) with the Ornstein-Uhlenbeck correlation function

$$\gamma(u) = \rho^{|u|}, \; 0 < \rho < 1 \qquad (3.1.1)$$

with equi-spaced design points on $[0,1]$ with spacing $1/n$. By defining $0^0 = 1$, we may include $\rho = 0$ in (3.1.1) to refer to the uncorrelated errors model. Recall that

$$\rho_n = \rho^{1/n} \qquad (3.1.2)$$

is therefore the adjacent correlation, that is, the correlation of neighboring errors (on the same subject) in model (2.1.1). These results will not require any distributional assumptions to be made for the error terms apart from the fact that the variance is finite.

The estimators to be considered are again of linear form

$$g_n(x) = \Sigma_{i=1}^{n} \; w_{ni}(x) \; \bar{y}(x_{ni}), \qquad (3.1.3)$$

which in matrix form is

$$g_n = W_n^{'} \; \bar{y}, \qquad (3.1.4)$$

where g_n is the vector of estimates when predicting at $x_{n1}, \ldots, x_{nn}$ and W_n is an (n by n) matrix whose $(i,j)^{th}$ element is $w_{ni}(x_{nj})$. When allowed to vary freely, the best possible linear (BL) choice for W_n, (with respect to minimizing the mean averaged squared error) is denoted

$$W_n^{BL} = [(\sigma^2/m)\Gamma_\rho + gg']^{-1} gg', \tag{3.1.5}$$

where Γ_ρ is the (n by n) correlation matrix with $(i,j)^{th}$ element $\rho_n^{|i-j|}$ and g is the n-vector of true values evaluated at the design points.

We will refer to the quantity

$$g_n^{BL}(x) = \Sigma_{i=1}^n w_{ni}^{BL}(x) \, \bar{y}(x_i) \tag{3.1.6}$$

as the best possible linear estimator of g. The matrix notation when estimating at the design points is

$$g_n^{BL} = [W_n^{BL}]' \, \bar{y}. \tag{3.1.7}$$

This is a slight abuse of the term estimator since g_n^{BL} depends on g; however, such usage is common in function estimation. We examined the limiting form of g_n^{BL} and the associated MASE and discovered that, unless $m \rightarrow \infty$ or $\rho = 0$, the best linear estimator fails to be consistent in our model. We set forth Theorem (2.8.1) which furnishes conditions on the model needed in order to obtain consistency.

In this chapter we will conduct a numerical study of four kernel estimators, whose names and primary references are given by

(1) PC: Priestley-Chao (Priestley and Chao, 1972),

(2) GM: Gasser-Müller (Gasser and Müller, 1979),

(3) NW: Nadaraya-Watson (Nadaraya (1964) and Watson(1964)), and

(4) GC: Gasser-Müller (cut-and-normalized) (Gasser and Müller, 1979).

We will refer to either (1) or (2) as a

NOCANE: not-cut-and-normalized estimator,

and to (3) or (4) as a

CANE: cut-and-normalized estimator.

This simply means that the weights in the estimator sum to one and hence the estimator is a true weighted average. The corresponding weights $w_{ni}(x)$ possess a functional form depending on x, x_{ni}, (and possibly $x_{n1}, \ldots, x_{nn}$), a bandwidth (smoothing) parameter h, and a given weighting (kernel) function $K(\cdot)$. In addition to the best linear estimator BL, we form estimators $g_n^{GC}(x)$, $g_n^{NW}(x)$, $g_n^{GM}(x)$, and $g_n^{GC}(x)$ in (3.1.3) by choosing $w_{ni}(x)$ of corresponding form:

(1) $w_{ni}^{PC}(x) = (1/h) \, (x_{n,i+1} - x_{ni}) \, K[(x - x_{ni})/h],$ (3.1.8)

     $i = 1, \ldots, n$ ($x_{n,n+1}$ arbitrarily defined);

(2) $w_{ni}^{NW}(x) = K[(x - x_{ni})/h] \, / \, \Sigma_t \, K[(x - x_{nt})/h],$ (3.1.9)

     $i = 1, \ldots, n;$

(3) $w_{ni}^{GM}(x) = (1/h) \int_{A_i} K[(x - u)/h] du,$ (3.1.10)

     $A_i = [s_{n,i-1}, s_{ni}], \quad s_{ni} = (x_{ni} + x_{n,i+1})/2;$

(4) $w_{ni}^{GC}(x) = w_{ni}^{GM}(x) \, / \, \Sigma_t \, w_{nt}^{GM}(x).$ (3.1.11)

In the customary treatment of these estimators, with $m = 1$ in an uncorrelated errors model, the bandwidth $h > 0$ depends on n and tends to zero at a suitable rate if n is allowed to tend to infinity. In our model, however, h will depend on m as well as n. When used in the context of model (2.1.1), h will be understood to be of form $h_{n,m}$. The kernels, which are typically probability densities, have support $[-1,1]$. Infinite support kernels are not considered since they smear boundary effects over the entire interval of estimation. We will return to retrictions on kernels later.

Provided g satisfies certain restrictions, we recall from Chapter 2 that the optimal kernel among convolution estimators

$$g_a(x) = \int_0^1 K(x,s) \, \bar{y}(s) \, ds \qquad (3.1.12)$$

in the Ornstein-Uhlenbeck model $(a = - \log \rho)$ is

$$K^*(x,s) = g(x)[a^2g(s)-g^{\cdot}(s)]\{(2a\sigma^2/m)+\int g(u)[a^2g(u)-g^{\cdot}(u)]du\}^{-1}. \qquad (3.1.13)$$

We note again that this kernel is of little use in practice since one does not know $g(x)$ or a in (3.1.13). However, we notice that $K^*(x,\cdot)$ depends on the correlation parameter a and is not necessarily non-negative. This suggests that possible improvement can be made in the kernel function by incorporating some estimate of the correlation. This could be a difficult problem, which will be a moot point if it turns out that using ordinary kernels and ignoring correlation performs nearly as well as the unrestricted linear estimator, whose limiting form led to the optimal kernel (3.1.13). One would further be interested in knowing the behavior of the best bandwidth, as a function of the amount of variance and correlation present, when the naive estimators PC, NW, GM, and GC are used and the covariance structure is taken into account. This may suggest that adjustments to the bandwidth may be of much greater importance than modification of the functional form of the kernel.

Toward these goals we have undertaken a small-scale numerical study, the details of which will be presented in Section 3.3. We first will review some analytic results, especially those of Hart and Wehrly (1986), who have studied model (2.1.1) in detail (for fairly general classes of correlation functions).

3.2 Recent Developments

Hart and Wehrly (1986) have considered model (2.1.1) under general conditions on the correlation function γ. It is implicit that γ is even and positive definite with

$$\gamma(0)=1, \quad |\gamma(u)| \leq 1 \text{ for all } u \text{ in } [-1,1]. \tag{3.2.1}$$

The estimator considered is the Gasser-Müller (GM) estimator specified by (3.1.10). For analysis of real data, the authors propose choosing h to minimize an estimate of the mean averaged squared error, which may be expressed as

$$M(h) = MASE[g_n^{GM}, g]$$

$$= (1/n)E\{\Sigma_{i=1}^n [g_n^{GM}(x_{ni}) - \bar{y}(x_{ni})]^2\} - (\sigma^2/m)[1 - (2/n)tr(W_h^{GM}\Gamma)]. \tag{3.2.2}$$

In the equispaced case the n distinct elements of the correlation matrix Γ are functions of

$$k\Delta = |x_{ni} - x_{n,i+k}|, \quad k=0,1,\ldots \tag{3.2.3}$$

so that one may estimate $\sigma^2\gamma(k\Delta)$ by

$$c(k) = (1/mn) \, \Sigma_{i=1}^m \Sigma_{j=1}^{n-k} [y_i(x_j) - \bar{y}(x_j)] [y_i(x_{j+k}) - \bar{y}(x_{j+k})]. \tag{3.2.4}$$

Hence $\hat{M}(h)$ is formed by dropping expectation in (3.2.2) and taking $\hat{\sigma}^2 = c(0)$ and $\hat{\gamma}(k\Delta) = c(k)/c(0)$. When using (3.2.2) and ignoring correlation, the authors report that the data may be either oversmoothed or undersmoothed, depending on the amount of correlation present and the relative sizes of σ^2 and n. Asymptotic expressions are given for the mean squared error and are summarized in the following theorems.

<u>Theorem (3.2.1)</u> -- Hart and Wehrly (1986)

Let x be in (0,1) and assume model (2.1.1) holds. Consider the GM estimator as specified by (3.1.10). Let us assume

(i) K has support $[-1,1]$ and is Lipschitz continuous of order 1, $\hspace{2cm}$ (3.2.5)

(ii) $|\gamma(u)-\gamma(v)| \leq B |u-v|$, some $B > 0$, and $\hspace{1cm}$ (3.2.6)

(iii) $\max_i |x_{ni}-x_{n,i-1}| = O(1/n)$. $\hspace{2cm}$ (3.2.7)

Then as n, m $\rightarrow \infty$, h $\rightarrow 0$,

(a) $\mathrm{Var}[g_n^{GM}(x)] = (\sigma^2/m)\int_{-1}^{1}\int_{-1}^{1} \gamma[(s-t)h]K(s)K(t)dsdt + O(1/nm)$. (3.2.8)

Suppose, in addition, we assume

(iv) g is Lipschitz continuous of order 1 on $[0,1]$. $\hspace{1cm}$ (3.2.9)

Then as n, m $\rightarrow \infty$ and h $\rightarrow 0$,

(b) $\mathrm{MSE}[g_n^{GM}(x)] = (\sigma^2/m) \int_{-1}^{1}\int_{-1}^{1} \gamma[(s-t)h] K(s)K(t)dsdt$

$\hspace{3cm} + [\int_{-1}^{1} g(x-hu) K(u)du - g(x)]^2 + o(1/n)$. $\hspace{1cm}$ (3.2.10)

<u>Theorem (3.2.2)</u> -- Hart and Wehrly (1986)

Suppose (i) - (iii) of Theorem (3.2.1) hold. Let $m/n = O(1)$ as n,m $\rightarrow \infty$ and assume that g is twice continuously differentiable on $[0,1]$ with $g''(x) \neq 0$. If γ has left and right hand derivatives at 0 with $\gamma'(0-) \neq \gamma'(0+)$, then as n,m $\rightarrow \infty$ and h $\rightarrow 0$,

$\hspace{1cm} \mathrm{MSE}[g_n^{GM}(x)] \sim (\sigma^2/m)[1 + \gamma'(0+)C_K h] + h^4\sigma_K^2[g''(x)/2]^2$, $\hspace{0.5cm}$ (3.2.11)

where

$$C_K = 2 \int_{-1}^{1}\int_{t}^{1} (s-t) K(s)K(t)dsdt \quad \text{and} \hspace{1cm} (3.2.12)$$

$$\sigma_K^2 = \int_{-1}^{1} u^2 K(u)du. \hspace{1cm} (3.2.13)$$

Furthermore, the bandwidth

$$h_m = [\sigma^2 \gamma'(0-)C_K / \sigma_K^4 (g'(x))^2]^{1/3} \, m^{-1/3} \qquad (3.2.14)$$

is optimum in the sense that, for all n and m sufficiently large,

$$MSE[g_n^{GM}(x;h_m)] < MSE[g_n^{GM}(x;h_{n,m})], \qquad (3.2.15)$$

where $h_{n,m}$ is any sequence of bandwidths tending to zero as $n,m \rightarrow \infty$ and such that

$$\lim_{\substack{n\rightarrow\infty, m\rightarrow\infty}} \inf \quad | \, h_m/h_{n,m} - 1 \, | > 0. \qquad (3.2.16)$$

<u>Theorem (3.2.3)</u> -- Hart and Wehrly (1986)

Let all the conditions of Theorem (3.2.2) hold with the exception that γ is assumed to be twice continuously differentiable on $[-1,1]$ with $\gamma''(0) \neq 0$. Then as $n,m \rightarrow \infty$ and $h \rightarrow 0$,

$$MSE[g_n^{GM}(x)] \sim (\sigma^2/m)[1 + \gamma''(0)\sigma_K^2 h^2] + h^4 \sigma_K^4 [g''(x)/2]^2. \quad (3.2.17)$$

Furthermore, if in addition $m/n = o(1)$, then the bandwidth

$$h_m^* = [-2\sigma^2 \gamma''(0) / \sigma_K^2 g''(x)^2]^{1/2} \, m^{-1/2} \qquad (3.2.18)$$

is optimum in the sense of Theorem (3.2.2).

The following observations are made:

(1) For m fixed in the GM estimator,

$$MSE[g_n^{GM}(x)] \rightarrow \sigma^2/m \text{ as } h \rightarrow 0, \, nh \rightarrow \infty, \quad (3.2.19)$$

which precludes consistency in this situation.

(2) If $m \rightarrow \infty$ in Theorem (3.2.1), we have mean square

consistency without requiring nh $\rightarrow$ 0. However,

Theorems (3.2.2) and (3.2.3) show that when m is not large

relative to n, nh $\rightarrow \infty$ is required in order to minimize

asymptotically the second order efficiency $MSE[g_n^{GM}(x)]$ -

(σ^2/m).

(3) Comparing h_m and h_m^* in (3.2.14) and (3.2.18), it is seen

that the less smooth the correlation function γ is at 0, the

larger the bandwidth. This makes sense because in this

situation, observations further away from the point of

estimation x will not be highly correlated with those errors

near x. Therefore a larger bandwidth does not simply obtain

redundant information about g(x).

The authors go on to demonstrate analytically the conditions

under which there would be a larger or smaller bandwidth than

expected with uncorrelated errors.

In another useful paper, Georgiev and Greblicki (1986) prove

that the PC, NW, and GM estimators satisfy the conditions of

Theorem (2.8.1) in the uncorrelated errors model with m = 1. In this

setting the following results for the PC, NW, and GM estimators are

proven.

Theorem (3.2.4) -- Georgiev and Greblicki (1986)

Let K be a probability density function such that

(i) K is almost everywhere continuous on R,   (3.2.20)

(ii) $|K(x)| \leq H(x)$, all x where H is symmetric,

    nonincreasing on $[0,\infty)$, and $\int H(y)dy < \infty$.   (3.2.21)

Let us further assume that

(iii) $\sup_x H(x) < \infty$,   (3.2.22)

(iv) $\max_i |x_{n,i} - x_{n,i-1}| = O(1/n)$,   (3.2.23)

(v) x is in (0,1) and is a continuity point of g, and   (3.2.24)

(vi) $h \to 0$, $nh \to \infty$ as $n \to \infty$.   (3.2.25)

Then, as $n \to \infty$,

$$E[g_n^{PC}(x) - g(x)]^2 \to 0, \text{ and} \qquad (3.2.26)$$

$$E[g_n^{GM}(x) - g(x)]^2 \to 0. \qquad (3.2.27)$$

If, in addition, we assume

(vii) $c \cdot I_{\{|x| \leq a\}} \leq K(x)$ for all x and some $c > 0$, $a > 0$, (3.2.28)

then

$$E[g_n^{NW}(x) - g(x)]^2 \to 0. \qquad (3.2.29)$$

    The condition (vi) is less restrictive than the conditions imposed on the PC estimate by Priestly and Chao (1972) and Cheng and Lin (1981a), who require the stronger assumption

$$nh^2 \to \infty \text{ as } n \to \infty. \qquad (3.2.30)$$

Also, unlike results of Gasser and Müller (1979), the result for the GM estimate requires neither a kernel with finite support nor continuity of g over the entire interval. The methods for these

proofs simply show the general conditions of Theorem (2.8.1) are satisfied for each $w_{ni}(x)$.

Other important work for the uncorrelated errors model includes the work of Gasser and Müller (1979,1984) and a series of papers by Georgiev (1984a,1984b,1984c,1984d) where the estimation of the derivatives $g^{(p)}(x)$, $p = 0,1,2,\ldots$ is treated.

3.3 Description of the Numerical Study

As mentioned in Section 3.1, the main questions to be answered in the numerical study may be summarized below. For a given n and various correlation-variance combinations,

(1) How well does BL perform? How well do PC, NW, GM, GC perform relative to BL?

(2) What is an approximation to the best bandwidth of each of the four estimators? How does it vary as a function of the variance and correlation? How is the optimum bandwidth affected if correlation is ignored? Does the expected value curve defined by $E[g_n(x;h^*)]$ seem to agree with the true $g(x)$ when the optimum bandwidth h^* is used?

(3) Which kernel estimators appear to do the best job? What is the effect of a cut-and-normalized estimator? Is there a need to modify the form of the estimator to take correlation into account?

A study of five functions defined over the range [0,1] was undertaken:

$$\text{(A)} \quad g_A(x) = \sin[\pi x + (\pi/2)] \qquad (3.3.1)$$

$$\text{(B)} \quad g_B(x) = 20 + \sin[\pi x + (\pi/2)] \qquad (3.3.2)$$

$$\text{(C)} \quad g_C(x) = 2 + 3x \qquad (3.3.3)$$

$$\text{(D)} \quad g_D(x) = 1 - x \qquad (3.3.4)$$

$$\text{(E)} \quad g_E(x) = \sin(2\pi x) \qquad (3.3.5)$$

These functions are all special cases of

$$g(x) = a + bx + \sin(cx + d)$$

for appropriate constants a, b, c and d. The FORTRAN program used in the numerical study was designed to handle any function of this type, the only restriction being that d must be 0 if c = 0. Integral expressions for certain theoretical quantities in Chapter 2 were worked out and calculated for arbitrary a, b, c and d in a subroutine. These quantities afforded a check on the reasonableness of some of the analytical results of Chapter 2 such as the limit of $MASE(W_n^{BL})$ and the limit of the boundary terms of $g'\Gamma^{-1}g$.

In selecting a, b, c and d for the study, it was decided to include a function (such as A, B) which possessed a vanishing first derivative at the boundaries 0 and 1. The corresponding effect on the boundary bias was of interest.

Function B is identical to A except for an additive constant of 20. It is desired to investigate whether this situation could lead to an optimal bandwidth different from function A. Intuition suggests that the same amount of smoothing should be done, although this turns out not to be the case for all estimators.

Functions C and D are straight line models chosen for comparison purposes. Function E is a pure sinusoid which vanishes at the endpoints. For this reason the CAN? would not necessarily be expected to outperform the NOCANE.

Functions (D) and (E) were considered by Benedetti (1974) for the PC and NW estimators in the uncorrelated model with $m = 1$. We will return to the reasons for our choice of functions in a later discussion.

The study was carried out for each function under a set of conditions corresponding to

(1) The Epanechnikov kernel [see Epanechnikov (1969)] in each

estimator: $K(u) = 0.75 (1-u^2) I_{(-1,1)}(u)$;   (3.3.6)

(2) $n = 10, 25$ for the number of design points

$x_{ni} = (i-0.5)/n, \ i=1,\ldots,n$;   (3.3.7)

(3) $\text{Var}[\bar{y}(x)] = \sigma^2/m = 10.0, 1.0, 0.1$ (3.3.8)

(This sequence corresponds to the effect of

increasing m, the number of subjects).

(4) $\rho = 0.0, 0.0001, 0.2000, 0.7000$ (3.3.9)

$\rho_{10} = 0.0, 0.3981, 0.8513, 0.9650$ (3.3.10)

$\rho_{25} = 0.0, 0.6918, 0.9377, 0.9858.$ (3.3.11)

We have $\rho_n = \rho^{1/n}$, as in (3.1.2). The spacing of design points is $1/n$ where (3.3.7) results in $s_{ni} = i/n$ for the Gasser-Müller integral bounds. Since m and σ^2 always appear together in the measure of error, we will regard σ^2/m as a single quantity representing the

common variance of the sample means $\bar{y}(x_{n1}), \ldots, \bar{y}(x_{nn})$.

Recall that ρ_n represents the correlation in adjacent within-subject errors. Suppose that the researcher believes (or estimates) that for n=10 equally-spaced design points, $\rho_n = 0.8$. By (3.1.2) we have the unit correlation $\rho = 0.1074$. If the researcher were then to have a finer grid of n=20 points available over the same x-range, the adjacent correlation would be $(0.1074)^{1/20} = 0.8944$, provided the errors follow the Ornstein-Uhlenbeck model. In our situation it is much easier to interpret $\rho_n = 0.8944$ than $\rho = 0.1074$ since the latter value represents the correlation at the extremes of the interval [0,1]. The use of ρ is convenient, however, as an input to a computer program designed to investigate the effect of finer spacings as we increase the value of n.

As a function of W_n, the mean averaged squared error may be equivalently written

$$\text{MASE}[W_n] = (\sigma^2/mn)\text{tr}(W_n' \Gamma W_n) + (1/n)(W_n'g - g)'(W_n'g - g), \quad (3.3.12)$$

where the terms represent the portions due to variance and squared bias. Efficient methods may be found for evaluating this expression and are outlined in the documentation of the computer routines available from the authors. For GM weights, Hart and Wehrly (1986) have also given a reduction of (3.3.12). Also recall that the BL MASE simplifies to

$$\text{MASE}[W_n^{BL}] = [g'g/n][1 - \text{tr}(W_n^{BL})]. \quad (3.3.13)$$

Now when using a kernel estimator, W_n is a matrix function of the design points, the bandwidth h, and the kernel function K. The following notation will be used when referring to a particular kernel

estimate:

Best Linear "Estimator": BL

Estimator: PC, NW, GM, GC

Bandwidth: (h; estimator)

Optimal bandwidth: $(h^*;$ estimator) $\hspace{2cm}$ (3.3.14)

Weight matrix: W(h;estimator)

Mean averaged squared error: M(h;estimator).

To obtain h^*, M(h;estimator) will need to be searched over some interval (0,a]. Since we are predicting on [0,1], a bandwidth larger that 1 will produce a very smooth estimator, and hence a smooth expected value of the estimator. A larger bandwidth than this, although possibly decreasing the MASE, will result in estimates with negligible visual differences. For this reason and other efficiency considerations, the bandwidth searches were truncated to the interval (0,2], which should be a large enough interval to assess the behavior of the estimate. M(h;estimator) was evaluated for an initial equi-spaced grid. The interval of uncertainty was then reduced to be the interval centered at the minimizing h by taking the new endpoints to be the neighboring values at which h was previously evaluated. On the next and subsequent passes there were two new evaluations made at points 1/4 and 3/4 of the distance between the endpoints of the interval of uncertainty. These were compared with the end and middle points already evaluated. The process was repeated until the width of the interval of uncertainty fell below a user-supplied tolerance, usually 0.02. Given the initial interval (0,2] and the above recipe

for searching, it is possible that the algorithm will choose exactly the same optimal h for more than one kernel estimator. If the minima were approximately the same, such results would not be surprising since there are a countable number of possible endings of evaluation sequences when the searches are conducted in the same manner.

To answer questions in (1), we may compare the difference of $M(h^*;$ estimator) and $M(W_n^{BL})$. The portions due to bias and variance follow from (3.3.12). The last two questions will be answered by analytic considerations in the next sections. Observe that since the kernel estimators are only bandwidth-modified for various ρ, the effect on the bandwidth of ignoring existing correlation may be obtained by comparing the $\rho = 0$ case to the desired $\rho > 0$ case. For cases with $\rho > 0$, however, we do not calculate $M(h_o;$estimator), where h_o is the best bandwidth when $\rho = 0$.

3.4 Results of the Numerical Study

Before discussing the empirical results, it will be useful to investigate the behavior of the estimators for large and small bandwidths (with n fixed) and how this might correspondingly affect the MASE criterion. Understanding this behavior will help explain certain aspects of the numerical study. We begin by observing the following properties of the (n by n) MASE matrix for each estimator, which we denote $W(h;$ estimator):

(1) For all h > 0, W(h;PC) and W(h;GM) are symmetric while

W(h;NW), W(h;GC) may not be symmetric. (3.4.1)

(2) For n fixed and h sufficiently small,

$$W(h;\text{estimator}) \approx c_h I_n, \qquad (3.4.2)$$

where c_h is a constant depending on the estimator.

Furthermore, if we let $h \to 0$,

$$c_h \to \infty \text{ for PC, and} \qquad (3.4.3)$$

$$c_h \to 1 \text{ for NW, GM, GC.} \qquad (3.4.4)$$

(3) As $h \to \infty$ (n fixed),

$$W(h;\text{estimator}) = d_h(1/n)(1_n 1_n') + o(1), \qquad (3.4.5)$$

where o(1) is a matrix with each element tending to 0. As

$h \to \infty$, the constant d_h depends upon the estimator used:

$$PC, GM: d_h \to 0, \text{ and} \qquad (3.4.6)$$

$$NW, GC: d_h \to 1. \qquad (3.4.7)$$

Now if a MASE matrix of form (3.4.2) is used, the corresponding MASE

becomes

$$M[c_h I_n] = c_h^2 (\sigma^2/m) + (c_h-1)^2(S_g^2 + \bar{g}^2), \qquad (3.4.8)$$

where we define

$$\bar{g} = \Sigma_{i=1}^n g(x_{ni})/n, \text{ and} \qquad (3.4.9)$$

$$S_g^2 = \Sigma_{i=1}^n [g(x_{ni}) - \bar{g}]^2/n. \qquad (3.4.10)$$

If we ignore the o(1) term and use a MASE matrix of form (3.4.5), the

form of the MASE is

$$M[d_h(1/n)1_n 1_n'] = d_h^2 \bar{\rho}_n(\sigma^2/m) + (d_h-1)^2 \bar{g}^2 + S_g^2, \qquad (3.4.11)$$

where $\bar{\rho}_n$ = the average of the elements of the correlation matrix Γ,

that is,

$$\bar{\rho}_n = (1/n^2) \; \Sigma_i \; \Sigma_j \; \rho_n^{|i-j|}. \qquad (3.4.12)$$

Observe that $0 < \bar{\rho}_n < 1$ and, for $\rho = 0$,

$$\bar{\rho}_n = 1/n \rightarrow 0 \text{ as } n \rightarrow \infty, \qquad (3.4.13)$$

while for $\rho > 0$,

$$\bar{\rho}_n \rightarrow \int_0^1 \int_0^1 \exp\{-\alpha|x-y|\} \; dx \; dy, \text{ as } n \rightarrow \infty. \qquad (3.4.14)$$

From equations (3.4.8) and (3.4.11) it is clear that for h large or small, the MASE for a given estimator depends on the relative sizes of $\mathrm{Var}[\bar{y}] = \sigma^2/m$, $\bar{g}$, and S_g^2.

As we shall soon see in the numerical results, there were situations where the optimum h was chosen extraordinarily large or small. Furthermore, there was often a large discrepancy in $M(h^*,\text{estimator})$ for the cut-and-normalized estimators NW, GC and the non-cut-and-normalized estimators PC and GM. In some situations, depending on the characteristics of the function (e.g., $\bar{g}$ and S_g^2), the NOCANE's had a smaller MASE than the CANE's. This behavior might be predictable whenever there is an extreme imbalance in the variance of the subject mean and the "variance" measure of the function g, that is, whenever

$$\mathrm{Var}[\bar{y}] \ll S_g^2 + \bar{g}^2 \quad \text{or} \quad \mathrm{Var}[\bar{y}] \gg S_g^2 + \bar{g}^2. \qquad (3.4.15)$$

These two competing measures of variation form the predominant components of the MASE. Whenever one is much larger than the other, the bandwidth for a given estimator will tend to be selected in a manner which makes c_h (or d_h) either very near 0 or very near 1, depending on the situation in (3.4.15). Examples of this behavior will be pointed out after the numerical summary is presented.

Summary tables of the computer study appear in the ensuing discussion. Tables (1a-1d) refer to function A, Tables (2a-2d) to function B, and so on. Tables labelled "a" and "c" contain the variable

$$M(h^*; \text{estimator}), \text{ for } n=10 \text{ and } n=25,$$

respectively. The best possible MASE for a linear estimator is also displayed in these tables. Tables labelled "b" and "d" contain the variable

$$(h^*; \text{estimator}), \text{ for } n=10 \text{ and } n=25,$$

respectively. Since there is little discrepancy between the results for $n=10$ and $n=25$, the reader may obtain the salient information on the optimal MASE and bandwidth by considering only tables labelled either "a" and "b" or tables labelled "c" and "d".

Comparisons of Functions A, B.

For this discussion we refer to Table 1a, Table 1b, Table 1c, and Table 1d for function A, and Table 2a, Table 2b, Table 2c, and Table 2d for function B.

For function A, $M(W^{BL})$ is smallest for low variance and high correlation, while for B it is smallest for low variance and low correlation. This is probably due to the fact that $\bar{g}_A^2 = 0$ and $\bar{g}_B^2 = 400.5$. However, there is little effect due to correlation when considering $M(W^{BL})$ for low variance, especially for function A. In these situations CANE's, with a moderate amount of smoothing, do significantly better than the NOCANE's and perform with ratios of about 3.0 in relation to the BL estimator.

Consider functions A and B with n = 10. For function A the worst case (0.39) for BL occurs for high variance and moderate correlation, while for B, the worst case (8.307) is for high variance and high correlation. Here we notice that, for A, NOCANE's with heavy smoothing do much better than the CANE's (which also use moderately heavy smoothing). For function B, the reverse is true. In this case, however, the NOCANE's prefer very low smoothing. There is little to choose between a NOCANE and a CANE in the worst case for B. We also point out that for almost all levels of $V(\bar{y})$ with function B, the correlation makes no difference in the performance for a NOCANE, that is, it depends almost entirely on the variance. When h is chosen smaller than 1/n, the correlation plays no part in the MASE.

Finally we observe that the CANE estimators perform almost the same for functions A and B. This is intuitively pleasing since we would expect the amount of smoothing to be invariant under additive shifts in the function. However, the smoothing of NOCANE's not only depends on the shift, but may result in a smaller measure of error than a CANE! From these results we see that the propensity to choose light or heavy smoothing for the CANE's depends on which of σ^2/m or $\bar{\rho}_n(\sigma^2/m) + S_g^2$ is the largest. In particular, if $\sigma^2/m < \bar{\rho}_n(\sigma^2/m) + S_g^2$ (which is equivalent to $\sigma^2/m < S_g^2/(1 - \bar{\rho}_n)$), then light smoothing is preferred. Now $S_g^2 = 0.5$ for both functions A and B. For n=10, $\bar{\rho}_n = 0.629$ when $\rho_n = 0.8513$. Hence, whenever $\sigma^2/m < 1.35$ and $\bar{\rho}_n = 0.629$, the CANE's prefer light smoothing. This is borne out in Table 3.4.1, since in this case a very large bandwidth for the CANE's is chosen only when $\sigma^2/m = 10 > 1.35$. We further observe that the

TABLE 1a. COMPARISONS OF OPTIMAL MASE: FUNCTION A, N=10
           VARIABLE = M(h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.3981	0.8513	0.9650
10	BL: BEST	0.333	0.386	0.390	0.279
	PC: NOCANE	0.614	0.758	1.301	1.640
	GM: NOCANE	0.614	0.758	1.301	1.639
	NW: CANE	1.356	2.539	6.762	9.275
	GC: CANE	1.356	2.539	6.762	9.274
1	BL: BEST	0.083	0.127	0.131	0.056
	PC: NOCANE	0.216	0.336	0.556	0.600
	GM: NOCANE	0.217	0.337	0.566	0.600
	NW: CANE	0.202	0.364	0.787	0.959
	GC: CANE	0.202	0.364	0.787	0.959
0.1	BL: BEST	0.010	0.016	0.017	0.006
	PC: NOCANE	0.051	0.072	0.099	0.106
	GM: NOCANE	0.051	0.070	0.093	0.100
	NW: CANE	0.030	0.049	0.087	0.098
	GC: CANE	0.030	0.049	0.087	0.098

TABLE 1b. COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION A, N=10
           VARIABLE = (h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.3981	0.8513	0.9650
10	BL: BEST	.	.	.	.
	PC: NOCANE	2.00	2.00	2.00	2.00
	GM: NOCANE	2.00	2.00	2.00	2.00
	NW: CANE	0.87	1.11	1.31	0.76
	GC: CANE	0.87	1.11	1.31	0.77
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.40	0.50	2.00	2.00
	GM: NOCANE	0.39	0.47	2.00	2.00
	NW: CANE	0.46	0.52	0.49	0.34
	GC: CANE	0.46	0.52	0.49	0.34
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.15	0.17	0.19	0.20
	GM: NOCANE	0.14	0.13	0.10	0.01
	NW: CANE	0.29	0.31	0.25	0.19
	GC: CANE	0.28	0.30	0.25	0.17

TABLE 1c. COMPARISONS OF OPTIMAL MASE: FUNCTION A, N=25
           VARIABLE = M(h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	0.222	0.374	0.388	0.278
	PC: NOCANE	0.508	0.740	1.296	1.638
	GM: NOCANE	0.508	0.740	1.296	1.637
	NW: CANE	0.642	2.397	6.729	9.267
	GC: CANE	0.642	2.397	6.729	9.267
1	BL: BEST	0.037	0.115	0.128	0.056
	PC: NOCANE	0.125	0.323	0.566	0.600
	GM: NOCANE	0.125	0.323	0.566	0.600
	NW: CANE	0.094	0.342	0.783	0.958
	GC: CANE	0.094	0.342	0.783	0.958
0.1	BL: BEST	0.004	0.014	0.017	0.006
	PC: NOCANE	0.033	0.070	0.084	0.084
	GM: NOCANE	0.033	0.070	0.097	0.100
	NW: CANE	0.014	0.046	0.087	0.098
	GC: CANE	0.014	0.046	0.087	0.098

TABLE 1d. COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION A, N=25
           VARIABLE = (h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	.	.	.	.
	PC: NOCANE	0.64	2.00	2.00	2.00
	GM: NOCANE	0.64	2.00	2.00	2.00
	NW: CANE	0.66	1.09	1.32	0.77
	GC: CANE	0.65	1.09	1.32	0.77
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.29	0.48	2.00	2.00
	GM: NOCANE	0.29	0.46	2.00	2.00
	NW: CANE	0.38	0.51	0.49	0.34
	GC: CANE	0.38	0.51	0.49	0.34
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.13	0.17	0.04	0.04
	GM: NOCANE	0.12	0.17	0.04	0.01
	NW: CANE	0.24	0.30	0.25	0.18
	GC: CANE	0.24	0.30	0.25	0.18

TABLE 2a. COMPARISONS OF OPTIMAL MASE: FUNCTION B, N=10
VARIABLE = M(h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION 0	0.3981	0.8513	0.9650
10	BL: BEST	0.998	2.041	5.706	8.307
	PC: NOCANE	7.331	9.082	9.799	9.799
	GM: NOCANE	7.107	10.000	10.000	10.000
	NW: CANE	1.356	2.539	6.762	9.275
	GC: CANE	1.356	2.539	6.762	9.274
1	BL: BEST	0.100	0.205	0.578	0.847
	PC: NOCANE	1.052	1.052	1.052	1.052
	GM: NOCANE	1.000	1.000	1.000	1.000
	NW: CANE	0.202	0.364	0.787	0.959
	GC: CANE	0.202	0.364	0.787	0.959
0.1	BL: BEST	0.010	0.021	0.058	0.085
	PC: NOCANE	0.177	0.177	0.177	0.177
	GM: NOCANE	0.100	0.100	0.100	0.100
	NW: CANE	0.030	0.049	0.087	0.098
	GC: CANE	0.030	0.049	0.087	0.098

TABLE 2b. COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION B, N=10
VARIABLE = (h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION 0	0.3981	0.8513	0.9650
10	BL: BEST	.	.	.	.
	PC: NOCANE	0.12	0.12	0.08	0.08
	GM: NOCANE	0.10	0.01	0.01	0.01
	NW: CANE	0.87	1.11	1.31	0.76
	GC: CANE	0.87	1.11	1.31	0.77
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.08	0.08	0.08	0.08
	GM: NOCANE	0.0.	0.01	0.01	0.01
	NW: CANE	0.46	0.52	0.49	0.34
	GC: CANE	0.46	0.52	0.49	0.34
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.08	0.08	0.08	0.08
	GM: NOCANE	0.01	0.01	0.01	0.01
	NW: CANE	0.29	0.31	0.25	0.19
	GC: CANE	0.28	0.30	0.25	0.17

67

```
TABLE 2c.   COMPARISONS OF OPTIMAL MASE: FUNCTION  B, N=25
            VARIABLE = M(h*;estimator)
```

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	0.400	1.852	5.549	8.233
	PC: NOCANE	6.004	10.075	11.816	12.179
	GM: NOCANE	5.625	10.000	10.000	10.000
	NW: CANE	0.642	2.397	6.729	9.267
	GC: CANE	0.642	2.397	6.729	9.267
1	BL: BEST	0.040	0.186	0.562	0.839
	PC: NOCANE	1.670	2.036	2.184	2.214
	GM: NOCANE	1.000	1.000	1.000	1.000
	NW: CANE	0.094	0.342	0.783	0.958
	GC: CANE	0.094	0.342	0.783	0.958
0.1	BL: BEST	0.004	0.019	0.056	0.084
	PC: NOCANE	1.260	1.296	1.311	1.314
	GM: NOCANE	0.100	0.100	0.100	0.100
	NW: CANE	0.014	0.046	0.087	0.098
	GC: CANE	0.014	0.046	0.087	0.098

```
TABLE 2d.   COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION  B, N=25
            VARIABLE = (h*;estimator)
```

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	.	.	.	.
	PC: NOCANE	0.07	0.07	0.07	0.07
	GM: NOCANE	0.05	0.01	0.01	0.01
	NW: CANE	0.66	1.09	1.32	0.77
	GC: CANE	0.65	1.09	1.32	0.77
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.05	0.05	0.05	0.05
	GM: NOCANE	0.01	0.01	0.01	0.01
	NW: CANE	0.38	0.51	0.49	0.34
	GC: CANE	0.38	0.51	0.49	0.34
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.05	0.05	0.05	0.05
	GM: NOCANE	0.01	0.01	0.01	0.01
	NW: CANE	0.24	0.30	0.25	0.18
	GC: CANE	0.24	0.30	0.25	0.18

jump from light to heavy smoothing can be rather dramatic as one goes from a small σ^2/m to a large σ^2/m. These results are basically the same for both sample sizes.

Comparisons of Functions C, D.

This discussion refers to Table 3a, Table 3b, Table 3c, and Table 3d for function C and Table 4a, Table 4b, Table 4c, and Table 4d for function D.

For function C, we observe that BL tends to be very good when both the variance and the correlation are low. There is a tendency toward very light smoothing in NOCANE's and moderate smoothing in CANE's, which perform slightly better than NOCANE's. The same is generally true for function D with the exception that the NOCANE's, with moderate smoothing, edge the CANE's (also under moderate smoothing).

The worst cases for BL are for high correlations and variances for both functions. Here the NOCANE's with moderate to very heavy smoothing outperform the CANE's and do reasonably well relative to BL, more so for function C than for function D. The CANE's also prefer moderate to very heavy smoothing. The performance of all estimators is more dependent on the variance than the correlation.

TABLE 3a.   COMPARISONS OF OPTIMAL MASE: FUNCTION   C, N=10
            VARIABLE = M(h*;estimator)

| VARIANCE | ESTIMATOR | ADJACENT CORRELATION | | | |
OF MEAN		0	0.3981	0.8513	0.9650
10	BL: BEST	0.928	1.748	3.621	3.465
	PC: NOCANE	1.926	2.786	4.863	5.826
	GM: NOCANE	1.932	2.786	4.864	5.826
	NW: CANE	1.452	2.688	6.931	9.343
	GC: CANE	1.452	2.688	6.931	9.342
1	BL: BEST	0.099	0.199	0.483	0.456
	PC: NOCANE	0.565	0.771	1.048	0.930
	GM: NOCANE	0.545	0.734	0.943	1.000
	NW: CANE	0.227	0.395	0.814	0.970
	GC: CANE	0.227	0.395	0.814	0.970
0.1	BL: BEST	0.010	0.020	0.050	0.047
	PC: NOCANE	0.100	0.100	0.100	0.100
	GM: NOCANE	0.100	0.100	0.100	0.100
	NW: CANE	0.037	0.058	0.092	0.099
	GC: CANE	0.037	0.058	0.092	0.099

TABLE 3b.   COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION   C, N=10
            VARIABLE = (h*;estimator)

| VARIANCE | ESTIMATOR | ADJACENT CORRELATION | | | |
OF MEAN		0	0.3981	0.8513	0.9650
10	BL: BEST	.	.	.	.
	PC: NOCANE	0.54	0.67	0.96	1.11
	GM: NOCANE	0.53	0.69	0.96	1.11
	NW: CANE	0.77	0.95	1.02	0.66
	GC: CANE	0.77	0.95	1.03	0.67
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.16	0.17	0.18	0.08
	GM: NOCANE	0.13	0.12	0.09	0.01
	NW: CANE	0.43	0.47	0.44	0.26
	GC: CANE	0.42	0.47	0.44	0.27
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.08	0.08	0.08	0.08
	GM: NOCANE	0.01	0.01	0.01	0.01
	NW: CANE	0.24	0.24	0.15	0.11
	GC: CANE	0.23	0.24	0.18	0.10

TABLE 3c. COMPARISONS OF OPTIMAL MASE: FUNCTION C, N=25
           VARIABLE = M(h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	0.388	1.598	3.513	3.363
	PC: NOCANE	1.245	2.705	4.858	5.829
	GM: NOCANE	1.246	2.705	4.858	5.829
	NW: CANE	0.693	2.544	6.903	9.336
	GC: CANE	0.693	2.544	6.903	9.336
1	BL: BEST	0.040	0.180	0.464	0.438
	PC: NOCANE	0.383	0.768	0.983	1.018
	GM: NOCANE	0.385	0.768	1.000	1.000
	NW: CANE	0.111	0.373	0.810	0.969
	GC: CANE	0.111	0.373	0.810	0.969
0.1	BL: BEST	0.004	0.018	0.048	0.045
	PC: NOCANE	0.089	0.167	0.185	0.188
	GM: NOCANE	0.100	0.100	0.100	0.100
	NW: CANE	0.018	0.054	0.091	0.099
	GC: CANE	0.018	0.054	0.091	0.099

TABLE 3d. COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION C, N=25
           VARIABLE = (h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	.	.	.	.
	PC: NOCANE	0.34	0.67	0.96	1.11
	GM: NOCANE	0.35	0.67	0.96	1.11
	NW: CANE	0.60	0.93	1.03	0.66
	GC: CANE	0.60	0.93	1.03	0.67
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.10	0.17	0.08	0.08
	GM: NOCANE	0.10	0.16	0.01	0.01
	NW: CANE	0.34	0.46	0.43	0.27
	GC: CANE	0.34	0.46	0.44	0.27
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.05	0.07	0.07	0.07
	GM: NOCANE	0.01	0.01	0.01	0.01
	NW: CANE	0.18	0.24	0.18	0.09
	GC: CANE	0.19	0.24	0.18	0.09

TABLE 4a. COMPARISONS OF OPTIMAL MASE: FUNCTION D, N=10
           VARIABLE = M(h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.3981	0.8513	0.9650
10	BL: BEST	0.249	0.283	0.303	0.283
	PC: NOCANE	0.312	0.456	0.999	1.337
	GM: NOCANE	0.312	0.456	0.999	1.337
	NW: CANE	1.077	2.191	6.390	9.006
	GC: CANE	1.077	2.191	6.390	9.006
1	BL: BEST	0.077	0.122	0.169	0.120
	PC: NOCANE	0.128	0.186	0.264	0.298
	GM: NOCANE	0.128	0.186	0.264	0.298
	NW: CANE	0.148	0.273	0.698	0.936
	GC: CANE	0.148	0.273	0.698	0.936
0.1	BL: BEST	0.010	0.018	0.031	0.018
	PC: NOCANE	0.037	0.053	0.085	0.090
	GM: NOCANE	0.037	0.053	0.086	0.095
	NW: CANE	0.023	0.040	0.082	0.097
	GC: CANE	0.023	0.040	0.082	0.097

TABLE 4b. COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION D, N=10
           VARIABLE = (h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.3981	0.8513	0.9650
10	BL: BEST	.	.	.	.
	PC: NOCANE	2.00	2.00	2.00	2.00
	GM: NOCANE	2.00	2.00	2.00	2.00
	NW: CANE	1.66	2.00	2.00	2.00
	GC: CANE	1.66	2.00	2.00	2.00
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.71	1.05	2.00	2.00
	GM: NOCANE	0.70	1.05	2.00	2.00
	NW: CANE	0.75	0.92	0.98	0.64
	GC: CANE	0.75	0.92	0.98	0.64
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.25	0.34	0.30	0.20
	GM: NOCANE	0.25	0.32	0.32	0.11
	NW: CANE	0.42	0.46	0.43	0.25
	GC: CANE	0.41	0.46	0.42	0.25

TABLE 4c. COMPARISONS OF OPTIMAL MASE: FUNCTION D, N=25
VARIABLE = M(h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	0.182	0.279	0.302	0.280
	PC: NOCANE	0.234	0.439	0.995	1.336
	GM: NOCANE	0.234	0.438	0.995	1.336
	NW: CANE	0.471	2.052	6.357	8.998
	GC: CANE	0.471	2.052	6.357	8.998
1	BL: BEST	0.036	0.112	0.164	0.116
	PC: NOCANE	0.080	0.181	0.264	0.298
	GM: NOCANE	0.080	0.181	0.264	0.298
	NW: CANE	0.071	0.259	0.695	0.935
	GC: CANE	0.071	0.259	0.695	0.935
0.1	BL: BEST	0.004	0.016	0.029	0.017
	PC: NOCANE	0.024	0.052	0.086	0.096
	GM: NOCANE	0.024	0.052	0.086	0.097
	NW: CANE	0.011	0.038	0.082	0.097
	GC: CANE	0.011	0.038	0.082	0.097

TABLE 4d. COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION D, N=25
VARIABLE = (h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	.	.	.	.
	PC: NOCANE	1.75	2.00	2.00	2.00
	GM: NOCANE	1.75	2.00	2.00	2.00
	NW: CANE	1.16	2.00	2.00	2.00
	GC: CANE	1.16	2.00	2.00	2.00
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.49	0.99	2.00	2.00
	GM: NOCANE	0.49	0.99	2.00	2.00
	NW: CANE	0.59	0.90	0.98	0.64
	GC: CANE	0.59	0.90	0.98	0.64
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.18	0.31	0.32	0.20
	GM: NOCANE	0.18	0.31	0.33	0.20
	NW: CANE	0.33	0.45	0.42	0.26
	GC: CANE	0.33	0.45	0.42	0.26

Results for Function E.

In this discussion we will refer to Table 5a Table 5b, Table 5c, and Table 5d for function E, an ordinary sinusoid.

For function E it is very interesting to note the similarity with the A-results for the BL MASE, which is identical when $\rho = 0$ over both sample sizes. We observe, however, that the NOCANE's are at least slightly better than CANE's for all combinations of variance and correlation. The performance of the NOCANE's and CANE's is expected to be comparable when we are considering a function for which $g(0) = g(1) = 0$.

3.5 Summary

In summary we make the following observations:

(1) Whether to cut-and-normalize depends not only on boundary conditions, but also considerations about the relative sizes of $V(\bar{y})$, ρ, $\bar{g}^2$ and S_g^2.

(2) For a given function, there are only minor differences between the two cut-and-normalized estimators and between the two NOCANE's. However, the difference between the two is not negligible asymptotically (see Gasser and Müller (1979)).

TABLE 5a. COMPARISONS OF OPTIMAL MASE: FUNCTION E, N=10
VARIABLE = M(h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.3981	0.8513	0.9650
10	BL: BEST	0.333	0.387	0.329	0.158
	PC: NOCANE	0.620	0.764	1.307	1.645
	GM: NOCANE	0.620	0.764	1.306	1.644
	NW: CANE	1.439	2.583	6.787	9.375
	GC: CANE	1.439	2.583	6.787	9.375
1	BL: BEST	0.083	0.127	0.080	0.022
	PC: NOCANE	0.223	0.358	0.572	0.606
	GM: NOCANE	0.222	0.362	0.572	0.606
	NW: CANE	0.290	0.471	0.877	0.985
	GC: CANE	0.290	0.470	0.877	0.985
0.1	BL: BEST	0.010	0.016	0.009	0.002
	PC: NOCANE	0.039	0.059	0.084	0.091
	GM: NOCANE	0.039	0.059	0.086	0.094
	NW: CANE	0.051	0.072	0.096	0.100
	GC: CANE	0.051	0.072	0.096	0.100

TABLE 5b. COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION E, N=10
VARIABLE = (h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.3981	0.8513	0.9650
10	BL: BEST	.	.	.	.
	PC: NOCANE	2.00	2.00	2.00	2.00
	GM: NOCANE	2.00	2.00	2.00	2.00
	NW: CANE	1.04	1.42	2.00	1.02
	GC: CANE	1.04	1.42	2.00	1.02
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.32	0.40	2.00	2.00
	GM: NOCANE	0.33	0.39	2.00	2.00
	NW: CANE	0.34	0.43	0.33	0.13
	GC: CANE	0.34	0.42	0.32	0.12
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.17	0.18	0.20	0.20
	GM: NOCANE	0.18	0.18	0.16	0.13
	NW: CANE	0.15	0.13	0.11	0.01
	GC: CANE	0.15	0.14	0.09	0.01

TABLE 5c. COMPARISONS OF OPTIMAL MASE: FUNCTION E, N=25
           VARIABLE = M(h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	0.222	0.377	0.312	0.141
	PC: NOCANE	0.542	0.746	1.302	1.643
	GM: NOCANE	0.542	0.746	1.302	1.643
	NW: CANE	0.783	2.446	6.755	9.372
	GC: CANE	0.783	2.446	6.755	9.372
1	BL: BEST	0.037	0.118	0.071	0.019
	PC: NOCANE	0.114	0.345	0.572	0.606
	GM: NOCANE	0.114	0.346	0.572	0.606
	NW: CANE	0.151	0.451	0.875	0.985
	GC: CANE	0.151	0.451	0.875	0.985
0.1	BL: BEST	0.004	0.015	0.008	0.002
	PC: NOCANE	0.019	0.055	0.084	0.092
	GM: NOCANE	0.020	0.055	0.085	0.093
	NW: CANE	0.027	0.068	0.096	0.100
	GC: CANE	0.026	0.068	0.096	0.100

TABLE 5d. COMPARISONS OF OPTIMAL BANDWIDTH: FUNCTION E, N=25
           VARIABLE = (h*;estimator)

VARIANCE OF MEAN	ESTIMATOR	ADJACENT CORRELATION			
		0	0.6918	0.9377	0.9858
10	BL: BEST	.	.	.	.
	PC: NOCANE	2.00	2.00	2.00	2.00
	GM: NOCANE	2.00	2.00	2.00	2.00
	NW: CANE	0.66	1.44	2.00	1.05
	GC: CANE	0.65	1.44	2.00	1.05
1	BL: BEST	.	.	.	.
	PC: NOCANE	0.26	0.36	2.00	2.00
	GM: NOCANE	0.26	0.38	2.00	2.00
	NW: CANE	0.25	0.39	0.30	0.13
	GC: CANE	0.25	0.39	0.30	0.13
0.1	BL: BEST	.	.	.	.
	PC: NOCANE	0.14	0.18	0.16	0.16
	GM: NOCANE	0.15	0.18	0.15	0.13
	NW: CANE	0.13	0.14	0.09	0.01
	GC: CANE	0.13	0.14	0.09	0.01

(3) The amount of smoothing for a given estimator generally increases with increasing correlation and variance. For a fixed variance, the requisite amount of smoothing tends to increase with correlation and decrease for larger ρ. This is consistent with the findings of Hart and Wehrly (1986). However, for a given ρ and $V(\bar{y})$, a CANE may choose significantly more or less smoothing than a NOCANE.

(4) As evidenced by functions A and B, the efficiency of a kernel estimator (when using MASE criterion) depends heavily on how $\bar{\rho}_n V(\bar{y})$ compares with $\bar{g}^2$ and S_g^2.

(5) If one ignores correlation when it actually is present, then, for a given estimator, there would probably be undersmoothing if the variance is high. If the variance is low then there can either be undersmoothing or oversmoothing.

(6) There seems to be a need to modify the estimator to be a function of the estimated correlation in certain situations where the BL estimate performs significantly better (e.g. function A with high correlation).

It appears that the presence of large correlation, especially with a large variance, tends to produce a very large bandwidth when using the MASE criterion. Large bandwidths tend to produce nearly flat, featureless estimates. Of course the variance in the estimator is decreased, but the bias may be rather large, especially near the boundaries. A very large positive correlation translates into sample paths very nearly parallel to the population regression function. If

the variance is small, the tendency would be toward a small bandwidth; a large variance would tend to call for over-smoothing. This is often the best a kernel estimate can do to correct for the fact that the estimate will tend to stay on one side of the regression function. For the most part, a flat (over-smoothed) estimate will stay on the same side of any continuous regression function g, but may at least cross g. Hence it seems natural for a kernel estimate to select, at the expense of the bias, a large bandwidth in these situations. It is conjectured that a variable bandwidth method (e.g. nearest neighbor estimation) may prove to be a more satisfactory estimate.

REFERENCES

Azzalini, A. (1984), "Estimation and Hypothesis Testing for
    Collections of Autoregressive Time Series," Biometrika, 71,
    85-90.

Benedetti, J. K. (1974), "Kernel Estimation of Regression
    Functions," Unpublished Doctoral Dissertation, University of
    Washington.

_____ (1975), "Kernel Estimation of Regression Functions,"
    Proceedings of the Computer Science and Statistics, ed. J. W.
    Frane, Eighth Annual Symposium on the Interface, February 13-14,
    1975, Health Science Computing Facility, UCLA, 405-412.

_____ (1977), "On the Nonparametric Estimation of Regression
    Functions," Journal of the Royal Statistical Society, Series B,
    39, 248-253.

Bishop, Y. M., Fienberg, S. E., and Holland, P. W. (1975), Discrete
    Multivariate Analysis: Theory and Practice, Cambridge: M. I.
    T. Press.

Cheng, K. and Lin, P. (1981a), "Nonparametric Estimation of a
    Regression Function," Zeitschrift fur
    Wahrscheinlichkeitstheorie und Verwandte Gebiete, 57, 223-233.

_____ (1981b), "Nonparametric Estimation of a Regression
    Function: Limiting Distribution," Australian Journal of
    Statistics, 23, 186-195.

Chung, K. L. (1974), A Course in Probability Theory, Second Edition,
    New York: Academic Press.

Clark, R. M. (1977), "Non-parametric Estimation of a Smooth
    Regression Function," Journal of the Royal Statistical Society,
    Series B, 39, 107-113.

Collomb, G. (1981), "Estimation Non-parametrique de la Regression:
    Revue Bibliographique," International Statistical Review, 49,
    75-93.

_____ (1985), "Nonparametric Regression: An Up-to-date
    Bibliography," Statistics, 16, 309-324.

Devroye, L. (1978), "The Uniform Convergence of the Nadaraya-Watson Regression Function Estimate," Canadian Journal of Statistics, 6, 2, 179-191.

Devroye, L. P., and Wagner, T. J. (1980a), "Distribution-Free Consistency Results in Nonparametric Discrimination and Regression Function Estimation," Annals of Statistics, 8, 2, 231-239.

_____ (1980b), "On the L_1 Convergence of Kernel Estimators of Regression with Applications in Discrimination," Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete, 51, 15-25.

Diggle, P. J., and Hutchinson, M. F. (1985), "Spline Smoothing with Autocorrelated Errors," A Progress Report ACT 85/14, CSIRO Division of Mathematics and Statistics.

Dym, H. and McKean, H. P. (1972), Fourier Series and Integrals: Probability and Mathematical Statistics, Volume 14, New York: Academic Press.

_____ (1976), Gaussian Processes, Function Theory, and the Inverse Spectral Problem, Probability and Mathematical Statistics, Volume 31, eds. Z. W. Birnbaum, E. Lukacs, New York: Academic Press.

Epanechnikov, V. A. (1969), "Nonparametric Estimates of a Multivariate Probability Density," Theory of Probability and Its Applications, 14, 153-158.

Eubank, R. L. (1986), "Nonparametric Regression Estimation," Unpublished Manuscript, Department of Statistics, Southern Methodist University.

Feller, W. (1971), An Introduction to Probability Theory and Its Applications, Volume II, Second Edition, New York: John Wiley and Sons.

Fulks, W. (1969), Advanced Calculus, Second Edition, New York: John Wiley and Sons.

Fuller, W. A. (1976), Introduction to Statistical Time Series, New York: John Wiley and Sons.

Gasser, T., and Müller, H. (1979), "Kernel Estimation of Regression Functions," In: Smoothing Techniques for Curve Estimation, eds. T. Gasser and M. Rosenblatt, LNM 757, 23-68.

_____ (1984), "Estimating Regression Functions and Their Derivatives by the Kernel Method," Scandinavian Journal of Statistics, 11, 171-185.

Gasser, T., Müller, H., Kohler, W., Molinari, L., and Prader, A. (1984), "Nonparametric Regression Analysis of Growth Curves," Annals of Statistics, 12, 1, 210-229.

Gasser, T., Müller, H., and Mammitzsch, V. (1985), "Kernels for Nonparametric Curve Estimation," Journal of the Royal Statistical Society, Series B, 47, 238-252.

Georgiev, A. A. (1984a), "Speed of Convergence Nonparametric Kernel Estimation of a Regression Function and Its Derivatives," Annals of the Institute of Statistical Mathematics, 36, A, 455-462.

_____ (1984b), "Nonparametric System Identification by Kernel Methods," IEEE Transactions on Automatic Control, 29, 4, 356-358.

_____ (1984c), "On the Recovery of Functions and Their Derivatives from Imperfect Measurements," IEEE Transactions on Systems Management Cybernetics, 14, 900-903.

_____ (1984d), "Kernel Estimates of Functions and Their Derivatives with Applications," Statistics & Probability Letters, 2, 45-50.

Georgiev, A. A. and Greblicki, W. (1986), "Nonparametric Function Recovering from Noisy Observations," Journal of Statistical Planning and Inference, 13, 1-14.

Glasbey, C. A. (1979), "Correlated Residuals in Non-linear Regression Applied to Growth Data," Applied Statistics, 28, 3, 251-259.

Graham, A. (1981), Kronecker Products and Matrix Calculus with Applications, Ellis Horwood Series in Mathematics and Its Applications, New York: John Wiley & Sons (Halsted Press).

Greblicki, W. and Krzyzak, A. (1980), "Asymptotic Properties of Kernel Estimators of a Regression Function," Journal of Statistical Planning and Inference, 4, 81-90.

Grenander, U. and Szego, G. (1958), Toeplitz Forms and Their Applications, Berkeley: University of California Press.

Grizzle, J. E., and Allen, D. M. (1969), "Analysis of Growth and Dose Response Curves," Biometrics, 25, 357-381.

Hardle, W. and Gasser, T. (1984), "Robust Non-parametric Function Fitting," _Journal of the Royal Statistical Society, Series B_, 46, 1, 42-51.

Hart, J. D. and Wehrly, T. E. (1986), "Kernel Regression Estimation Using Repeated Measurements Data," To appear: _Journal of the American Statistical Association_.

Huber, P. J. (1981), _Robust Statistics_, New York: John Wiley and Sons.

Johnston, G. J. (1982), "Probabilities of Maximal Deviations for Nonparametric Regression Function Estimates," _Journal of Multivariate Analysis_, 12, 402-414.

Mack, Y. P., and Silverman, B. W. (1982), "Weak and Strong Uniform Consistency of Kernel Regression Estimates," _Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete_, 61, 405-415.

Morrison, D. F. (1970), "The Optimal Spacing of Repeated Measurements," _Biometrics_, 26, 281-290.

Muirhead, R. J. (1982), _Aspects of Multivariate Statistical Theory_, New York: John Wiley and Sons.

Nadaraya, E. A. (1964), "On Estimating Regression," _Theory of Probability and Its Applications_, 9, 141-142.

_____ (1970), "Remarks on Non-parametric Estimates for Density Functions and Regression Curves," _Theory of Probability and Its Applications_, 15, 134-137.

_____ (1973), "Some Limit Theorems Related to Nonparametric Estimates of Regression Curve," _Bulletin of Academy of Science Georgian S.S.R._, 71, 57-60.

_____ (1983a), "Limit Distribution of the Maximum Deviation of Estimates of a Regression Curve," _Trudy Tbiliss. University_, 239, 43-64.

_____ (1983b), "A Limit Distribution of the Square Error Deviation of Nonparametric Estimators of the Regression Function," _Zeitschrift fur Wahrscheinlichkeitstheorie und Verwandte Gebiete_, 64, 1, 37-48.

Noda, K. (1976), "Estimation of a Regression Function by the Parzen Kernel-type Density Estimators," _Annals of the Institute of Statistical Mathematics_, 28, 221-234.

Parzen, E. (1960), *Modern Probability Theory and Its Applications*,
     New York: John Wiley and Sons.

_____ (1962a), "On Estimation of a Probability Density Function
     and Mode," *Annals of Mathematical Statistics*, 33, 1065-1076.

_____ (1962b), *Stochastic Processes*, San Francisco: Holden-Day.

_____ (1967), *Time Series Analysis Papers*, eds.: E. Parzen and
     G. M. Jenkins, San Francisco: Holden-Day.

Prakasa Rao, B. L. S. (1983), *Nonparametric Functional Estimation*,
     New York: Academic Press.

Priestley, M. B., and Chao, M. T. (1972), "Non-parametric Function
     Fitting," *Journal of the Royal Statistical Society, Series B*,
     34, 385-392.

Rao, C. R. (1973), *Linear Statistical Inference and Its Applications,
     Second Edition*, New York: John Wiley and Sons.

Rosenblatt, M. (1956), "Remarks on Some Nonparametric Estimates of a
     Density Function," *Annals of Mathematical Statistics*, 27,
     832-837.

_____ (1969), "Conditional Probability Density and Regression
     Estimators," In: *Multivariate Analysis II*, ed. P. R. Krishnaiah
     25-31, New York: Academic Press.

_____ (1971), "Curve Estimates," *Annals of Mathematical
     Statistics*, 42, 6, 1815-1842.

Schuster, E. F. (1972), "Estimation of a Probability Density
     Function and Its Derivatives," *Annals of Statistics*, 40,
     1187-1195.

Schuster, E. F. and Yakowitz, S. (1979), "Contributions to the
     Theory of Non-parametric Regression, with Application to System
     Identification," *Annals of Statistics*, 7, 1, 139-149.

Serfling, R. J. (1980), *Approximation Theorems of Mathematical
     Statistics*, New York: John Wiley and Sons.

Spiegelman, C., and Sacks, J. (1980), "Consistent Window Estimation
     in Nonparametric Regression," *Annals of Statistics*, 8, 2,
     240-246.

Stone, C. J. (1977), "Consistent Nonparametric Regression," *Annals
     of Statistics*, 5, 4, 595-645.

Tapia, R. A. and Thompson, J. R. (1978), Nonparametric Probability Density Estimation, Baltimore: Johns Hopkins University Press.

Watson, G. S. (1964), "Smooth Regression Analysis," Sankhya, Series A, 26, 359-372.

Whittaker, E. T. and Watson, G. N. (1963), A Course in Modern Analysis, Fourth Edition, United Kingdom: Cambridge University Press.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 3	**2. GOVT ACCESSION NO.**	**3. RECIPIENT'S CATALOG NUMBER**
4. TITLE (and Subtitle) Considerations for the Linear Estimation of a Regression Function When the Data are Correlated		**5. TYPE OF REPORT & PERIOD COVERED** Technical
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) D. B. Holiday, T. E. Wehrly, and Jeffrey D. Hart		**8. CONTRACT OR GRANT NUMBER(s)** N00014-85-K-0723
9. PERFORMING ORGANIZATION NAME AND ADDRESS Texas A&M University Department of Statistics College Station, TX 77843		**10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS**
11. CONTROLLING OFFICE NAME AND ADDRESS		**12. REPORT DATE** June 1987
		13. NUMBER OF PAGES 86
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		**15. SECURITY CLASS. (of this report)** Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

NA

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Nonparametric regression; growth curves; correlated data; optimum bandwidth; mean averaged squared error.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A repeated-measurements model, applicable in growth curve analysis, with correlated errors within subjects is developed. Kernel estimators of the population regression function are examined for various correlation functions. Limiting forms of an optimal linear combination of the subject means derived. Conditions for consistency of a general linear estimator are stated for the Ornstein-Uhlenbeck correlation function and a more general covariance structure. A numerical study investigating the requisite amount of smoothing and the efficiency of four popular kernel estimators is carried out. The expected values of the estimators are compared against one another and against an optimal linear combination.

DD FORM 1473 **1 JAN 73** EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

END

8-87

DTIC